From the Aging Research Center, Department
of Neurobiology, Care Sciences and Society
Karolinska Institutet, Stockholm, Sweden

DOPAMINE, DECISION-MAKING, AND AGING

NEURAL AND BEHAVIOURAL CORRELATES

Lieke de Boer

Karolinska Institutet

Stockholm 2019

Cover illustration: *Fractals* by Kike Jonkeren, inspired by the stimulus material for the cognitive tasks used in the research for this thesis.

Published by Karolinska Institutet.
Printed by Arkitektkopia AB, 2019
ISBN 978-91-7831-551-2

Dopamine, decision-making, and aging

Neural and behavioural correlates

THESIS FOR DOCTORAL DEGREE (Ph.D.)

The thesis will be defended at Samuelssonsalen, Tomtebodavägen 6, Karolinska Institutet, Solna

Thursday, October 3, 2019, at 10:00 am

By

Lieke de Boer

Principal supervisor:

Dr Marc Guitart-Masip
Aging Research Center
Department of Neurobiology,
Care Sciences and Society
Karolinska Institutet

Co-supervisors:

Professor Lars Bäckman
Karolinska Institutet
Department of Neurobiology,
Care Sciences and Society
Aging Research Center

Professor Lars Nyberg
Umeå University
Umeå Center for
Functional Brain Imaging
Department of Radiation Sciences

Opponent:

Dr Hanneke den Ouden
Donders Institute for Brain, Cognition
and Behaviour
Radboud University

Examination board:

Professor Tomas Furmark
Uppsala University
Department for Emotion Psychology

Dr Johan Lundström
Karolinska Institutet
Department of Clinical Neuroscience

Dr Signe Allerup Vangkilde
University of Copenhagen
Department of Psychology

ABSTRACT

On any given day, we make tons of decisions. These can be as simple as deciding how to dress or what to eat, or more complex, such as whether to spend or invest money. Good decision-making involves being able to select the best alternative from a range of options, and adjust one's preferences based on what is happening in the environment. As humans get older, their ability to do this changes. Age-related changes in decision-making ability result from changes in brain structure and function, such as the deterioration of the brain's dopaminergic system in old age. In this thesis, we used a sample of 30 older and 30 younger participants to investigate age-related differences in neural and behavioural correlates of value-based decision-making, which involves making decisions that can result in rewards and punishments. Such decisions are known to rely on dopaminergic functioning. In the brain, we have looked at neural activity reflecting value and reward prediction errors (RPEs), the availability of dopamine D1 receptors, and integrity of white matter microstructure. For the behavioural data, we have used computational modelling to disentangle motivational biases and other parameters reflecting parts of the learning process that underlies successful decision-making.

In **study 1**, we investigated whether performance on a value-based decision-making task differed between the two age groups. We also looked at whether performance differences could be explained by differential neural processing of RPEs and expected value in the striatum and prefrontal cortex (PFC). We used a novel computational model to estimate expected value, decision uncertainty and confidence. We found that older adults earned fewer rewards on the task. The number of rewards earned could be predicted by the strength of the neural signal reflecting expected value in the ventromedial PFC (vmPFC), which was attenuated in older adults. Beyond age, the strength of this neural signal could be predicted by dopamine D1 receptor (D1-R) availability in the nucleus accumbens (NAcc). In **study 2**, we showed that integrity of white matter microstructure in the pathway between the NAcc and vmPFC also predicted the neural value signal in the vmPFC, independently of age and D1-R availability in the NAcc.

In **study 3** and **4**, we focused on dissociating the effects of action and valence on neural and behavioural correlates of decision-making. In **study 3**, we used computational modelling to characterize faster learning to act in response to rewards, and abstaining from acting in response to punishments, as being the result of biased instrumental learning. Study 3 also showed that variability in dopamine D1-R availability could be divided into cortical, dorsal striatal and ventral striatal components. Regardless of age, dopamine D1-R availability in the dorsal striatal component was related to biased learning from rewarded actions. In **study 4** we investigated anticipatory value signals after learning had reached an asymptote.

We observed no differences between age groups in anticipatory neural responses to action and valence, and no relationship between anticipatory neural signals and dopamine D1-R availability. Older adults did show an attenuated punishment prediction error signal in the insula, compared with younger adults. The strength of differentiation between reward- and punishment prediction error signals in the insula was related to dopamine D1-R availability in the cortex.

These studies have demonstrated that the existing theoretical framework surrounding the role of dopamine system in decision-making and aging fits with dopamine D1-R availability data and behavioural data in older and younger adults, and partly explain why older adults show behavioural differences in value-based decision-making tasks. Collectively, the studies in this thesis provide important multimodal evidence that increases our understanding of the neural correlates that underlie value-based decision-making and how they are affected in healthy aging.

LIST OF SCIENTIFIC PAPERS

I. **de Boer, L.**, Axelsson J., Riklund, K., Nyberg, L., Dayan, P., Bäckman, L. and Guitart-Masip, M. (2017). Attenuation of Dopamine-Modulated Prefrontal Value Signals Underlies Probabilistic Reward Learning Deficits in Old Age. *eLife* 6 (September). doi:10.7554/eLife.26424.

II. **de Boer, L.**, Garzón, B., Axelsson J., Riklund, K., Nyberg, L., Bäckman, L. and Guitart-Masip, M. Corticostriatal white matter integrity and dopamine D1 receptor availability independently predict age differences in prefrontal value signaling during reward learning. *Under Review*.

III. **de Boer, L.**, Axelsson J., Riklund, K., Nyberg, L., Bäckman, L. and Guitart-Masip, M. (2019). Dorsal striatal dopamine D1 receptor availability predicts an instrumental bias in action learning. *Proceedings of the National Academy of Sciences.* 116(1), 261-270. doi: 10.1073/pnas.1816704116.

IV. **de Boer, L.**, Axelsson J., Riklund, K., Nyberg, L., Dayan, P., Bäckman, L. and Guitart-Masip, M. The representation of Go/NoGo patterns and dopamine D1 receptor availability. *Manuscript.*

LIST OF RELATED SCIENTIFIC PAPERS NOT INCLUDED IN THE THESIS

I. **de Boer, L.**, Plavén-Sigray, P., Betts, M.J., Düzel, E., Guitart-Masip, M., Decomposing the neural value signal: A registered report. *Under review*

II. Perosa, V. **de Boer, L.**, Guitart-Masip, M., Apostolova, I. Buchert, R. Ziegler, G., Metzger, C., Amthauer, H. Düzel, E., Betts, M.J. The role of the striatum in learning to orthogonalize action and valence: a combined PET and 7T MRI aging study. *Under review*

III. Betts, M.J., Richter, A., **de Boer, L.**, Tegelbeckers, J., Perosa, V., Chowdhury, R., Dolan, R.J., Seidenbecher, C., Schott, B.H., Düzel, E., Guitart-Masip, M., Krauel, K. Learning in anticipation of reward and punishment: Perspectives across the human lifespan. *Under review*. preprint available at https://www.biorxiv.org/content/10.1101/738211v1

IV. Adams, R.A., Moutoussis, M., Nour, M.M., Dahoun, T., Lewis, D., Illingworth, B., Veronese, M., Mathys, C., **de Boer, L.**, Guitart-Masip, M., Friston, K.J., Howes, O.D., Roiser, J.P. Comparing active inference and reinforcement learning models of human decision-making and striatal dopamine signalling. *Under review*

V. L Ricciardi, P Haggard, **L de Boer**, C Sorbera, MP Stenner, F Morgante, MJ Edwards. (2017). Acting without being in control: Exploring volition in Parkinson's disease with impulsive compulsive behaviours. *Parkinsonism & related disorders* 40, 51-57

VI. E Loh, M Deacon, **L de Boer**, RJ Dolan, E Duzel (2016). Sharing a context with other rewarding events increases the probability that neutral events will be recollected. *Frontiers in human neuroscience* 9, 683

CONTENTS

Introduction 1
- Computational models of decision-making 2
  - Reinforcement learning 2
  - Rescorla-Wagner 3
  - TD-learning 4
  - Bayesian observer models 5
  - After value estimation 5
  - Extensions of action propensities 7
  - Model definition 7
- Dopamine and the basal ganglia in value-based learning 8
- Basal ganglia – cortex interface 14
- Dopamine in the cortex 16
- Dopamine, decision-making and aging 17
- How can we study the dopaminergic system? 18

Aim of the thesis 21

Methods 22
- Research participants and procedure 22
- Task descriptions 23
- Computational modelling 26
- Two-armed bandit – Rescorla-Wagner (study 1) 26
- Two-armed bandit – Bayesian observer model (study 1 and 2) 27
- Go/No-Go – Rescorla-Wagner (study 3) 29
- Model fitting 31
- Model selection 32
- PET imaging 32
- PET image acquisition 33
- PET analysis 33
- Definition of regions of interest for PET 34
- Principal Component Analysis 34
- Magnetic resonance imaging (MRI) 34
- Diffusion-weighted imaging (DWI) 35
- DWI preprocessing and analysis 35
- fMRI 36
- fMRI acquisition 36
- fMRI analysis 37
- Statistical analyses 38
- Exclusion and analysis samples for different studies 39
- Statistical disclosure statement 40
- Data and code availability 41

Individual studies 42
Study 1 42
Study 2 43
Study 3 45
Study 4 47
Discussion 50
Acknowledgements 60
References 62

LIST OF ABBREVIATIONS

5-HT	serotonin
ACC	anterior cingulate cortex
ANOVA	analysis of variance
BIC	Bayesian Information Criterion
BP_{ND}	non-displaceable binding
CT	computed tomography
D1-R	dopamine type 1 receptor
D2-R	dopamine type 2 receptor
dACC	dorsal ACC
dlPFC	dorsolateral PFC
DTI	diffusion tensor imaging
DWI	diffusion weighted imaging
fMRI	functional MRI
FSL	FMRIB software library
FWHM	full-width half-maximum
GABA	gamma-aminobutyric acid
GL	go to avoid losing
GLM	general linear model
GPe	globus pallidus pars externa
GPi	globus pallidus pars interna
GW	go to win
L-DOPA	levodopa
lOFC	lateral OFC
MANOVA	multivariate ANOVA
mOFC	medial OFC
MRI	magnetic resonance imaging
MSN	medium spiny neuron
NAcc	nucleus accumbens
NGL	no-go to avoid losing
NGW	no-go to win
NMDA	N-Methyl-d-aspartate
OFC	orbitofrontal cortex

PCA	principal component analysis
PD	Parkinson's Disease
PET	positron emission tomography
PFC	prefrontal cortex
RL	reinforcement learning
ROI	region of interest
RPE	reward prediction error
RT	reaction time
SN	substantia nigra
SNr	SN pars retriculata
SPM	statistical parametric mapping
STN	subthalamic nucleus
TAB	two-armed bandit
TD	temporal difference
vmPFC	ventromedial PFC
VS	ventral striatum
VTA	ventral tegmental area
WAIC	watanabe-akaike information criterion

INTRODUCTION

As humans age, many changes occur in the brain. Among these are changes in neurochemical integrity. For example, the number of cells producing and receptors responding to the neurotransmitter dopamine, decrease at a rate of between 5-10% per decade from the age of 30 (Kaasinen et al., 2000; Suhara et al., 1991; Volkow et al., 1994, 1998). Aging also affects the brain's structural integrity. Both longitudinal and cross-sectional studies have demonstrated that the total grey matter, which contains the cell bodies of neurons, decreases across the lifespan. Similarly, both the microstructural integrity (Moscufo et al., 2018) and the total amount of white matter, containing myelinated axons that connect neurons, decreases across the lifespan (Grady, 2012).

Because the brain is important for cognition, it is not surprising that these large changes in brain structure are paralleled by a decrease in a number of cognitive abilities. For example, older people have lower processing speed (Salthouse, 1992), consistently perform worse on tasks that measure working memory (Gazzaley et al., 2005; Salthouse, 1992; West, 1999), episodic memory (Old and Naveh-Benjamin, 2008; Rönnlund et al., 2005), cognitive flexibility (Chao and Knight, 1997; Lindenberger and Baltes, 1997) and decision-making (Eppinger et al., 2011; Mell et al., 2005). The simultaneous decline in structural brain integrity and cognitive performance means that aging provides a natural experimental setting to investigate how the brain works, allowing for the investigation of correlations between cognition, age, and brain structure and function, including dopaminergic integrity (Bäckman et al., 2006).

In this thesis, I will focus on dopamine and decision-making. Decision-making is a broad term, within psychology defined as "*the cognitive process resulting in the selection of a belief or a course of action among several alternative possibilities*" (Wang and Ruhe, 2007). Decision-makers have to weight evidence reflecting benefits associated with several options, and select one of the options in light of that evidence (Montague et al., 2012). Specifically, we use value-based decision-making tasks. These are experimental situations where decisions are made based on the learnt expected values of a set of choices that lead to rewards or punishments (Daw and Doya, 2006). Older adults show suboptimal performance on value-based decision-making tasks in experimental settings compared with younger adults, especially when they have to adapt their behaviour to changing environments (Mell et al., 2005).

In the remainder of this introduction, I will discuss these interrelated topics. First, I will discuss how value-based decision-making can be studied with computational models. Second, I will give a brief overview of the neuroanatomy and the role of the dopaminergic system in the human brain, and its role in value-based decision-

making. Third, I will review evidence on how aging interacts with the dopaminergic system and how effects of aging modulate decision-making performance. Fourth, I will discuss the ways in which we can study the dopaminergic system in humans, and why we used positron emission tomography (PET) in the studies in this thesis.

Computational models of decision-making

Decision-making in the natural environment is hugely complex, and may at first seem difficult to formally investigate. However, the development of computational modelling techniques has made it possible for researchers to study intricate decision-making processes. These models can be set up in many different ways, but will always include a number of parameters that determine how individuals learn about which choices are good, and which choices are bad. These parameters take on different values for each individual, which results in different predicted preferences for each person. The purpose of these individualised models is for them to closely mimic each individual's behaviour. Once these parameter values have been determined, they can be correlated with features of that individual's neural integrity, such as markers of the dopaminergic system, brain structure or brain activity in task-relevant brain areas, with the goal of increasing our understanding of how patterns of brain activation, as well as indicators of neurotransmitter functioning, are involved in behaviour. I will outline a few different approaches to computational modelling below.

Reinforcement learning

Reinforcement learning (RL) is a term from the field of computer science, and refers to how learners incorporate reward-related information about the environment into their behaviour in order to maximize future rewards. RL algorithms formally describe this learning process (Sutton and Barto, 1998). In this thesis, we use RL algorithms to create agents who imitate a human performing a value-based decision-making task. The modelled RL agent iteratively estimates values for each available choice at each task trial, and makes decisions based on those values. The values are estimated and updated when the agent receives rewards (and punishments) as a consequence of their choices. The agent will then adapt their behaviour, usually adhering to the policy of maximising the number (or magnitude) of positive reinforcements and minimising the number (or magnitude) of negative reinforcements.

A more detailed RL description of a value-based decision-making task goes something like this: At each time step, agents who are in a certain *state* select an *action* from a set of possible actions. This action leads to a *reward*, which can be positive (reward) or negative (punishment). The type and magnitude of the reward as a consequence of the selected action is determined by the *environment*.

In the studies presented in this thesis, the environment is determined by the task contingencies that map choices to outcomes. When the reward is experienced, the state of the agent is updated and the cycle is repeated. Agents in RL problems learn which states are most rewarding, and how good it is to perform a given action, given a certain state. The agent then builds up a set of learnt expected values about different alternatives in the environment. The rules by which this occurs are described by a *value function* (Sutton and Barto, 1998).

Rescorla-Wagner

A simple value function often used in computational models of decision-making is the Rescorla-Wagner learning rule. This rule is strictly speaking not taken from RL in the computer science context, but was instead developed as a result of classical conditioning experiments, where animals associatively learned about stimuli from positive and negative reinforcements (Daw and Tobler, 2013; Rescorla and Wagner, 1972). The Rescorla-Wagner rule is derived from the idea that the discrepancy between a received and expected reward as the result of a conditioned stimulus informs future expectations about that stimulus. The difference between the received and expected reward is called *prediction error,* and the learning rule updates the value of the conditioned stimulus based on that prediction error.

Formally, the Rescorla-Wagner rule approximates the value (V) of a stimulus *s* on trial t+1 as follows:

$$V_{t+1}(s_t) = V_t(s_t) + \alpha \cdot \delta_t \quad (1)$$

In English, this formula states that the value V of a stimulus *s* on trial *t+1* is equal to the value of that stimulus on trial *t,* with the addition of the prediction error δ. The prediction error is multiplied by a learning rate α (with $0 \leq \alpha \leq 1$), that determines to what extent the future value estimate is affected by the prediction error. As stated above, the prediction error is defined as the difference between expected and obtained value on a given choice:

$$\delta_t = r_t - V_t(s_t) \quad (2)$$

In stable environments, the value estimate for choice *a* converges on the mean probability of reward for that stimulus, because a stable environment leads to predictable reward rates, reducing δ. In such a stable environment, the learning rate indicates how fast the value converges onto this mean value. The learning rate may be slow in such situations, because occasional unexpected rewards should not affect expected values in these environments too much. In more unstable environments, reward feedback will be more variable, and prediction errors will be greater. Learning should also be faster, as changes in reward feedback may reflect actual changes in the environment. Therefore, α may change with increased uncertainty about the environment (Behrens et al., 2007; Dayan et al., 2000).

The Rescorla-Wagner learning rule could be conceived of as a simple value function in RL: an organism updates the value of a stimulus (or action, or state) based on the difference between expectations and reality. Because of the way dopamine neurons respond to rewarding and non-rewarding outcomes, Rescorla-Wagner models are often used when studying the dopaminergic system (Bayer and Glimcher, 2005; Chowdhury et al., 2013a; Daw, 2011; Daw et al., 2006; Guitart-Masip et al., 2011, 2014a; Hamid et al., 2016; Pessiglione et al., 2006; Swart et al., 2017).

The Rescorla-Wagner learning rule cannot fully account for the temporal dynamics of dopaminergic firing in the midbrain (Schultz et al., 1997). Other more elaborate algorithms from the field of RL have been useful in capturing some crucial aspects of neural and behavioural aspects of decision-making. First, whereas the Rescorla-Wagner rule is "trial-based" (Schultz et al., 1997), other RL algorithms distinguish between states, actions and outcomes, which means that stimulus onset at the beginning of a trial can carry value. Second, the Rescorla-Wagner rule does not take into account the goal of the decision-maker (Daw and Tobler, 2013). This is not surprising, given that it was developed as a rule to describe classical conditioning. However, decision-makers have the long-term goal of maximising rewards. This is another problem that can be solved by other RL algorithms. Here, I will specifically discuss Q-learning, a form of temporal difference (TD) learning (Sutton and Barto, 1998).

TD-learning

As stated above, one of the crucial differences between the Rescorla-Wagner learning rule and TD is that TD also considers stimulus presentations as carrying information about future reward. This is because a stimulus could signal a state transition to an individual, and the new state could predict more future rewards. The TD control algorithm that resembles the Rescorla-Wagner rule most is known as Q-learning. The important insight that TD algorithms such as Q-learning provided was captured by the Bellman Equation, which states that the value of any state *s* is a long sum over a series of expected future discounted rewards that can be obtained from that state (Daw and Tobler, 2013). If the average predictions of the Bellman Equation are faulty, the agent will observe a prediction error, and the sum of future expectations is updated in a very similar way to the Rescorla-Wagner learning rule.

Q-learning is very useful in situations where a sequence of choices leads to an eventual reward. In the studies in this thesis, the choices that are made almost always lead to immediate rewards, which allows for immediate updating of choice values. In addition, there is often no final goal to the tasks we use. Therefore, the expected future rewards in the decision-making processes that we model are approximated on a trial-by-trial basis by the Rescorla-Wagner rule. However, for

our analyses, we borrow the TD-learning assumption that states have values, so that stimulus presentation at the trial onset evokes a prediction error. Another element that we have borrowed from TD-learning is notation: it is conventional in RL to use Q to mean state-action values (the expected value Q that an agent has for choice a), whereas V reflects simple state values (the expected value V that an agent has for stimulus i). Q therefore becomes action-specific, whereas V is stimulus-specific.

Bayesian observer models

Another approach to defining a value function is the use of Bayesian observer models (Daunizeau et al., 2010; Payzan-LeNestour and Bossaerts, 2011). Bayesian observer models are strictly speaking not within the RL paradigm, but can also provide estimates of values. Whereas models of the RL family estimate single value estimates for each option in the environment, Bayesian observer models provide a probability distribution around value estimates. The Bayesian observer is a theoretical entity that performs a task, taking into account the information about the environment, using Bayes' rule to estimate what the posterior representation of the environment based on priors and the likelihood of obtaining the received reward under that prior. Bayesian observers track the probability distribution over expected values of outcomes, incorporating the notion of uncertainty about the environment, and weighting the observations about reward probability by the level of uncertainty about the environment. Bayesian models can be *hierarchical:* they allow for tracking not only of mean and variance on a trial by trial level, but also the dynamics of this process, i.e. they can incorporate the rate of change in variability, or *volatility* of the environment.

After value estimation

In the computational models we use, agents perform actions. In doing so, they need to balance *exploitation* and *exploration* (Badre et al., 2012; Cohen et al., 2007; Dayan and Sejnowski, 1996; Frank et al., 2009; Wilson et al., 2014). Exploitation refers to choosing the alternative for which the agent has computed the largest expected value. If the environment is stable, the same choice will repeatedly lead to the greatest reward, and minimal exploration is needed. In practice however, environments are volatile, and in order for exploitation to return the greatest rewards, the agent needs information about the value of all available action-state contingencies. Thus, the agent needs to explore the available actions.

Regardless of what value function is used, the balance between exploitation and exploration is often formalised with a softmax function, which defines the probability that an agent will explore or exploit by translating the estimated value difference between the two into a choice probability function in favour of the

most valuable option, with some leeway for exploration depending on temperature parameter β (Figure 1). The softmax is an *observational model* or *response model,* a model that computes how value estimations lead to observed choices. In the interest of brevity, I will only discuss the softmax as an observational model, as it is very commonly used in RL (Daw, 2011; Sutton and Barto, 1998).

The softmax shows that the probability that an agent chooses option 1 on the estimated value (V) for that choice, relative the action values of other choices (option 2 in Figure 1). In a situation where an agent is faced with two choices, the softmax rule favours the action with the highest action value, and assigns the largest choice probability to this action. The other options has a chance of being explored proportional to how valuable they are perceived to be. In the softmax above, β is an *inverse temperature parameter,* which allows for a larger or smaller differentiation between choice probabilities. β is usually individually fitted. If β is very large (as in the figure where $\beta = 100$), the probability of choosing the action with the highest action value will be relatively much greater than the probability of choosing the other alternatives. If β is very small (as in the figure where $\beta = 1$), the probability of all actions becomes nearly equiprobable, even with large value differences. Thus, β can be seen as an indicator of how much the agent's decision is based on value. This can reflect a tendency to explore (if β is low) or low interest in the task or in obtaining the highest valued choice.

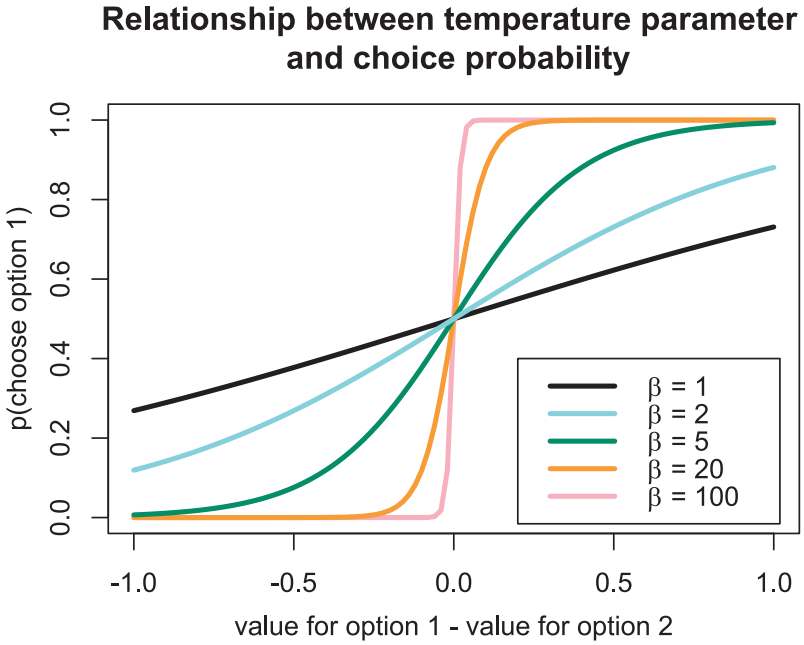

Figure 1. *When the inverse temperature parameter β is large, small differences between options lead to deterministic behaviour. When β is small, the agent may explore the other less valuable options more.*

Extensions of action propensities

When applying RL to cognitive neuroscience, one needs to define the learning process that is believed to have given rise to the observed data. In order to do this, one first decides which parameters are added to the value function. This more elaborate definition of the value function is called the *learning model* (Daw, 2011). The learning model will reflect hypotheses about the algorithms that the brain uses to solve RL problems (Daw, 2011). When such parameters are added, the resulting action value is referred to as an action propensity. An example of a simple learning model is one where action propensities m are defined as $m_a(t) = Q_a(t)$. In this case, the learning model is reduced to the standard RW rule or TD algorithm as presented in equation 1. Usually however, models are expanded with additional parameters that affect action propensities, because the goal of defining a learning model is to understand how agents learn and to understand their behaviour. Because humans do not always behave rationally, these value functions do not simply capture rational approaches to decision-making: often parameters are included that reflect common biases that make certain actions more attractive to agents, despite these biases leading to suboptimal performance.

Examples of such additions include perseveration parameters, which reflect the common observation that people often tend to stick with the same choice, or avoid choice repetition, regardless of the calculated value difference between choices (Laskowski et al., 2016; Rutledge et al., 2009). Additionally, previous studies have included parameters that reflected different sensitivities to rewards and punishments (Guitart-Masip et al., 2014a; Kobayakawa et al., 2010), a *go bias,* for Go/NoGo paradigms, as well as a parameters reflecting *motivational biases,* which captures the tendency of individuals to pair actions with rewards and inactions with avoiding punishments (Cavanagh et al., 2013; Guitart-Masip et al., 2012a, 2014a; Huys et al., 2011, 2012; Swart et al., 2017). When researchers are interested in studying the exploration/exploitation balance, a parameter reflecting an exploration bonus may be included in the model definition (Dayan and Sejnowski, 1996; Frank et al., 2009). Similarly, if risk-taking is of interest to behaviour, a model parameter reflecting risk preference is commonly added to the value function (Niv et al., 2012; Rouault et al., 2019).

Model definition

When deciding which learning model may have given rise the observed data, there are two ways in which the definition of the learning model needs to be specified. One is *parameter estimation*, the numerical estimation of different free parameters that go into the definition of the model, e.g. the value of the learning rate, forgetting rate, temperature or bias parameters. Note that the model here includes both the learning model and the observational model, as the temperature parameters are part of the observational model. These values are estimated by

calculating how likely the data would have been observed under different parameter values, giving rise to a *likelihood,* quantifying how well parameters fit the observed data. There are different approaches to parameter estimation. One approach is to perform a *grid search*, which requires trying out every possible combination of parameter values within a specified range. Grid searches will in practice result in the best fitting parameter values, but are computationally hugely taxing and very slow. For this reason, the most common approaches to parameter estimation involve *gradient descent*. This approach will transverse the multidimensional parameter space and calculate the log-likelihood for a combination of parameter values and its neighbours, and then travel in the direction where the gradient towards the lowest log-likelihood is steepest. In order to dictate which parameter values to consider, one can use *priors*, which put more prior probability on the range of values within the prior distribution.

The other way in which the model needs to be defined is the parameters to include in the model, e.g. the decision of whether a forgetting rate should be included at all. This process is done by performing *model comparison* between different learning models. Model comparison is done by looking at all the likelihoods for different agents, and choosing the model that gives rise to the best likelihood overall, penalizing for the number of parameters to prevent overfitting using *information criteria* (Daw, 2011) such as the Bayesian information criterion (BIC) or Watanabe–Akaike information criterion (WAIC).

Dopamine and the basal ganglia in value-based learning

As described above, Rescorla-Wagner and TD models within the computational framework of RL are often used to study value-based decision-making, because of how dopamine neurons in the midbrain respond to unexpected rewards and non-rewards. Dopamine's function as a neurotransmitter was discovered in the 1950s and soon after linked to symptoms of Parkinson's disease (PD), a disease characterised by bradykinesia (Carlsson, 1959; Marsden, 2006). Research has since implicated the dopaminergic system in a range of cognitive and motor processes. These processes include control and invigoration of motor functions (Beierholm et al., 2013; Robertson et al., 2015), facilitation of high-level cognitive processes like working memory and cognitive flexibility (Cools and D'Esposito, 2011; Gruber et al., 2006; Sawaguchi and Goldman-Rakic, 1991; Takahashi et al., 2012), responsiveness to reward (Aarts et al., 2012; Schultz, 1998; Schultz et al., 1997), and learning from reward prediction errors (RPEs) (Bayer and Glimcher, 2005; Hart et al., 2014; Pessiglione et al., 2006).

Value-based decision-making became a popular way of studying dopamine after Schultz et al., 1997 discovered with neurophysiology recordings in monkeys that unexpected rewards lead to a burst in activity in dopamine neurons located in

the midbrain. The midbrain contains the ventral tegmental area (VTA) and the substantia nigra (SN). Schultz et al. (1997) observed that when a reward was expected, the reward-predicting stimulus caused a burst in activity in dopamine neurons at the time of this stimulus presentation (Cohen et al., 2012; Pan et al., 2005; Schultz et al., 1997), but not at the time of reward delivery. The burst in activity in dopamine neurons after a predictive stimulus is presented is thought of as a TD RPE, as the animal is put from a neutral state in a rewarding state (Schultz, 1998). It is also thought to reflect an action invigoration signal, encouraging or discouraging the animal to respond to this stimulus (Hamid et al., 2016).

Unexpected rewards, on the other hand, cause a burst of activity in dopamine neurons at the time of reward delivery, and unexpected reward omissions cause a dip in activity in dopamine neurons at the time of omission (Schultz et al., 1997). These bursts scale proportionally to the size of rewards (Tobler et al., 2005). The burst in dopamine neuron activity in response to unexpected rewards are commonly thought to represent RPEs at the time of reward delivery, resembling the computational prediction error δ as described in the RL section above. Since, many studies confirmed the existence of RPEs in the SN/VTA with neurophysiology recordings (Bayer and Glimcher, 2005; Cohen et al., 2012), although a debate about the ability of dopamine neurons to effectively signal negative RPEs exist (Bayer et al., 2007; Fiorillo, 2013).

Neurons in SN/VTA releasing dopamine in response to RPEs project to different target areas in the brain, which has an effect on associative learning. One major target is the striatum, consisting of caudate, putamen and nucleus accumbens (NAcc). The striatum is often roughly divided in a ventromedial to dorsolateral gradient, with the dorsal striatum connecting to premotor and motor cortical areas, and the ventral striatum (VS) connecting to cortical areas that mediate reward and emotional processing (figure 3; increasingly dark > increasingly ventromedial, (Haber and Knutson, 2010)). There are three major pathways through which dopamine operates. First, the nigrostriatal pathway includes dopamine neurons in the SN pars compacta (SNc) that project to the dorsal parts of the striatum, involved in the control of movement. In PD, the nigrostriatal pathway is most affected. The other two pathways are thought to be involved in cognition, executive control and value-based decision-making: one is the mesocortical pathway, and connects the VTA to the prefrontal cortex (PFC) directly. The other is termed the mesolimbic pathway, and connects the VTA to the VS. Neural RPEs are thought to be expressed in the VS through the mesolimbic pathway (Arias-Carrión et al., 2010).

In the striatum, medium spiny neurons (MSNs) express dopamine receptors and are responsive to dopamine release. dopamine binds to two families of receptor subtypes in MSN terminals: D1-like receptors (which the D1 and D5 receptor type belong to, from now on referred to as D1 receptors, D1-Rs) and D2-like receptors (which the D2, D3 and D4 receptor type belong to, from now on referred to as

D2 receptors, D2-Rs). D1 and D2-Rs respond to dopamine differently. Generally speaking, the binding of dopamine to D1-Rs has an excitatory effect on the post-synaptic cell, increasing the likelihood that the neuron will fire. Conversely, when dopamine binds to D2-Rs, this has an inhibitory effect on the cell, reducing the likelihood of the cell to fire (Clark and White, 1987; Keeler et al., 2014). MSNs with D1 and D2 receptors are approximately evenly expressed in striatum, but segregated in different populations (Keeler et al., 2014).

D1-Rs are expressed by MSNs that form the direct pathway. D2-Rs are expressed by MSNs that form the indirect pathway (Keeler et al., 2014). Both pathways affect activity in thalamocortical projections, which are by default tonically inhibited, as a result of inhibitory projections from the internal globus pallidus (Figure 2). Activation of the direct pathway activates inhibitory projections to the internal globus pallidus. This decreases the tonically inhibitory pressure on the thalamus, leading to a facilitation of connections with the cortex. Activation of the indirect pathway via the subthalamic nucleus (STN) leads to activation of the internal globus pallidus and thalamus, which increases inhibitory firing of the thalamus, ultimately decreasing excitatory feedback to cerebral cortex (Arias-Carrión et al., 2010) and inhibiting connections with the cerebral cortex. D1 activation increases activation in the direct pathway, whereas D2 activation decreases activation in the indirect pathway (Frank et al., 2004). D1 and D2 receptors are differentially responsive to tonic and phasic changes in dopamine. D2-Rs have high affinity to dopamine, and are therefore sensitive even to small and transient, or tonic changes in dopamine levels. In contrast, D1-Rs have low affinity, and are therefore most sensitive to large changes in dopamine levels (Marcott et al., 2014; Surmeier et al., 2011). Dopamine release from SN/VTA in response to positive RPEs is phasic and therefore primarily acts on D1-Rs and direct pathway, facilitating corticostriatal connections, whereas dopamine dips from SN/VTA in response to negative RPEs are tonic and act on D2-Rs and increase activity in the indirect pathway (Hikida et al., 2010).

In the dorsal striatum, activity of dopamine neurons resulting from expectation of reward have the effect of increasing the likelihood of acting, as the direct pathway facilitates action selection (Wickens et al., 2003). However, the temporal specificity of dopamine release resulting from better than expected outcomes (RPEs) also leads to plasticity in the direct pathway, which has the effect of increasing the likelihood that the action that directly predated the unexpected reward is repeated. Conversely, a dip in dopamine results in the activation of the indirect pathway and the inhibition of motor output. Additionally, dips resulting from unexpected reward omissions or punishments leads to plasticity in the indirect pathway, which decreases the likelihood of repeating an action that led to reward omission or punishment next time the same circumstances are encountered (Collins and Frank, 2014).

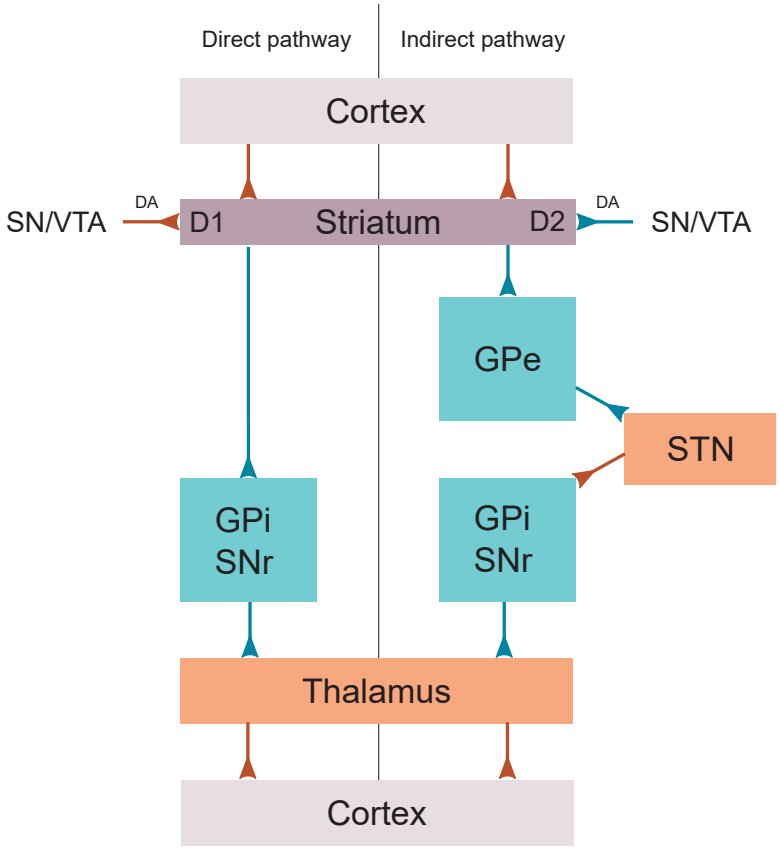

Figure 2. *Schematic representation of the direct and indirect pathway in the striatum. Red lines between areas represent excitatory connections, blue lines represent inhibitory connections. Glutamatergic neurons from the cortex project to MSNs in the striatum. These MSNs are also the target of dopaminergic projections from the SN/VTA in the midbrain. When dopamine binds to D1-Rs which form the direct pathway, the effect of this activation on postsynaptic neurons is excitatory. Thus, the GABA-ergic projections from the striatum to the GPi SNR are activated, which decreases the overall inhibitory pressure on the thalamus, and facilitates activation of the cortex. Activation of D2-Rs leads to an inhibitory effect on the postsynaptic cell, indirectly increasing the inhibitory pressure on the thalamus, implying that activation of both pathways leads to increased connectivity between the thalamus and the cortex. Figure adapted from Leisman et al., (2013).*

This functional organisation of the circuit has inspired neural network models of learning to repeat actions that lead to reward and to inhibit actions that lead to punishment. For example, Frank et al. (2004), proposed a neural network model inspired by this architecture of the striatum and its modulation by dopamine that is able to perform various RL tasks (Frank et al., 2004). This model is largely supported by research findings, although the picture is not always consistent. For example, the effects of dopaminergic enhancement sometimes affects only reward learning (Beierholm et al., 2013; Guitart-Masip et al., 2012b; Pessiglione et al., 2006; Rutledge et al., 2009), and sometimes both reward and punishment learning (Bódi et al., 2009; Cools et al., 2009; Guitart-Masip et al., 2014a;

Moustafa et al., 2008). The variability in findings are sometimes attributed to baseline dopamine levels (Cools et al., 2009; Cox et al., 2015; Swart et al., 2017) or genetic modulatory factors (Frank et al., 2009; den Ouden et al., 2013). Other more elaborate models of the role of the two pathways in value-based decision-making and action selection have been proposed by for example (Hwang, 2013; Keeler et al., 2014))

Although the direct/indirect pathway models are based on the findings on the dorsal striatum, where the direct pathway facilitates action selection, and the indirect pathway facilitates inhibitions of actions (Wickens, 1990; Wickens et al., 2003), this functional organization of the striatum is conserved along the ventromedial to dorsolateral axis, with the exception of the NAcc (Kupchik et al., 2015), from which MSNs containing D1-Rs also project to the GPe. The striatum is interfaced with the cortex in a limbic-cognitive-motor gradient, with the most ventromedial areas being most responsive to value-related and cognitive signals elsewhere in the brain, and the most dorsolateral area receiving mostly input from premotor and motor areas (Haber and Behrens, 2014, Figure 3). This has led to the idea that the striatum functions as a selection device, promoting the selection of actions in the dorsal striatum, but the selection of cognitive processes, goals and mental representations in more ventromedial regions.

In recent decades, the RPE signal has been extensively researched in humans using functional magnetic resonance imaging (fMRI) and value-based decision-making tasks that implement RL as a computational framework. RPEs have been found to be represented in VS (Niv et al., 2012; Pagnoni et al., 2002; Pessiglione et al., 2006), an area in the brain to which the mesolimbic midbrain projects. However, not all fMRI studies demonstrate a signal in VS positively correlating with reward, and negatively correlating with expected value, as one would expect from a canonical RPE signal (Chowdhury et al., 2013a; Wimmer et al., 2014).

In human neuroimaging studies, neural signals in SN/VTA and VS in humans show a pattern of response incompatible with pure RPE signalling, as these signals were predominantly predicted by action requirements. RPEs in both the SN/VTA and VS humans were only observed when rewards were the result of actions, and not when rewards where the result of omitting an action (Guitart-Masip et al., 2011). This observation is suggestive of a Pavlovian mechanism, which couples activation with appetitive cues, and inhibition with aversive cues (Guitart-Masip et al., 2014b; Haber and Behrens, 2014). The opposite (coupling activation with aversive cues, and inhibition with appetitive cues) is much more difficult to learn (Boureau and Dayan, 2011; Guitart-Masip et al., 2012a). This mechanism is reminiscent of the functional anatomy of the striatum described above. The direct and indirect pathway couples actions and reward reinforcement with approach behaviour through dopamine release in response to reward

and reward-predicting cues. This reinforces the direct pathway, facilitating the learning and repetition of the recently executed action. Conversely, it couples inaction with the avoidance of punishment, by lowering dopamine release in response to punishment (and reward omission) and punishment-predicting cues. This reinforces the indirect pathway, preventing the repetition of the recently executed action and promoting delearning of the association between stimulus and action selection (Frank et al., 2004; Soares-Cunha et al., 2016). Recent studies have shown that the instrumental learning of actions in response to rewards is biased towards the coupling of action and valence (Chowdhury et al., 2013b; Guitart-Masip et al., 2012a), and that manipulation of the dopaminergic system can affect this bias (Guitart-Masip et al., 2014a; Swart et al., 2017), which is in line with this mechanism.

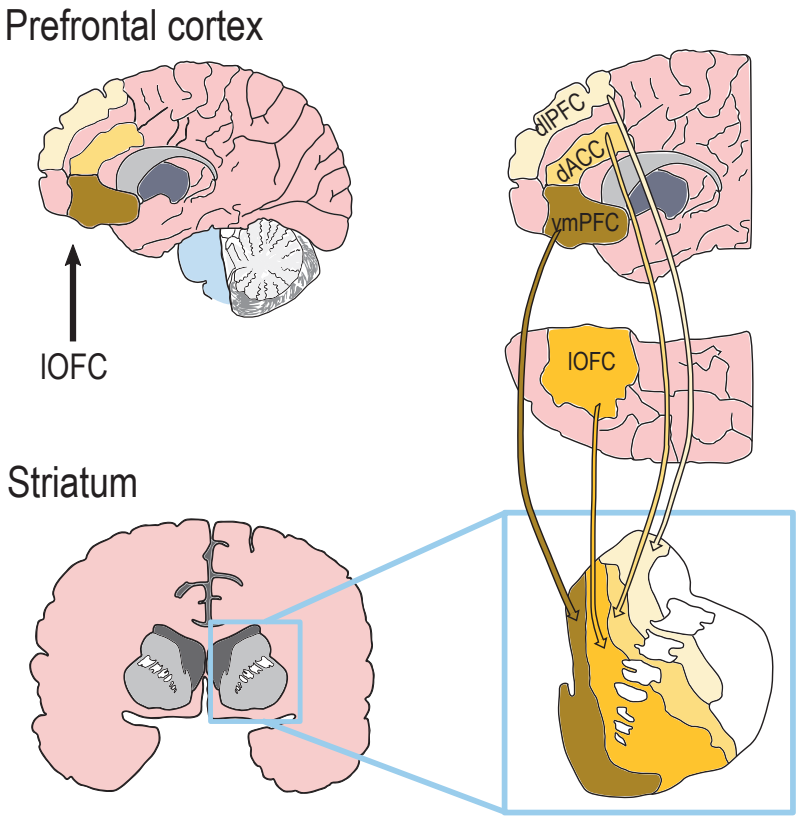

Figure 3. *Schematic representation of prefrontal projections to the striatum. The top left image shows a sagittal plane of the brain, with the dlPFC, dACC and vmPFC coloured from light to dark yellow. The striatum is represented here in the middle of the brain in grey. The lOFC is located on the lateral surface of the prefrontal cortex. dlPFC, dACC, lOFC, and vmPFC are all part of the corticothalamostriatal loop complex. They project in that order from dorsolateral to increasingly ventromedial areas in the striatum (bottom right), where their projections converge. Activity in corticostriatal loops regulates reward processing and action selection. Adapted from Haber and Knutson, (2010).*

It is important to consider that dopamine does not act alone in the learning and reinforcing of actions leading to rewards or punishments. For example, the neurotransmitter serotonin (5-HT) has been suggested to be important for the inhibition of behaviour that leads to punishment and increased sensitivity to aversive outcomes (Boureau and Dayan, 2011; Cools et al., 2011; Daw et al., 2002; Geurts et al., 2013; den Ouden et al., 2013), forming an opponency axis with dopamine. There is also a range of evidence implicating 5-HT in punishment learning and risky choice (Rogers, 2011). Additionally, the neurotransmitter acetylcholine projects to many of the same structures as dopamine (Fobbs and Mizumori, 2014), and is thought to be important for many different types of cognition, including attention and memory (Picciotto et al., 2012). Noradrenaline is another abundant neurotransmitter that plays important roles in arousal, motivation and attention (Nieuwenhuis et al., 2005). NA has also been suggested to signal an uncertainty prediction error, in a similar way that dopamine signals RPEs (Preuschoff et al., 2011).

Basal ganglia – cortex interface

Cortical signals mandate the activity in the direct and indirect pathways in striatum, through roughly topographical glutamatergic cortical projections to the striatum (Haber and Behrens, 2014; Haber and Knutson, 2010, Figure 3). These afferent connections are modulated by dopamine in the striatum (Seamans and Yang, 2004).

These projections, together with the basal ganglia circuit, form cortico-striato-nigro-thalamo-cortical loops (Haber and Knutson, 2010). Functionally different cortical areas project to different subregions of striatum, with functionally similar cortical areas projecting to similar parts of the striatum. Initially the idea emerged that several broadly functionally segregated and separate cortical loops existed, broadly divisible into limbic, associative and sensorimotor circuits. However, the idea that these loops are integrated recently received much support, based on the idea that in order for goal-directed action and adaptive behaviour to occur, reward evaluation, associative learning and developing motor execution plans are required (Haber and Knutson, 2010).

This cortical input includes a range of regions which will be discussed in this section. Areas relevant to value-based decision-making include the ventromedial PFC and medial orbitofrontal cortex (vmPFC/mOFC), lateral OFC (lOFC, figure 3) and insula (Haber and Behrens, 2014; Rushworth et al., 2011). Examples of additional areas important for decision-making are dorsolateral PFC (dlPFC) and dorsal anterior cingulate cortex (dACC), and frontopolar cortex (FPC) which will not be discussed in as much detail here.

In humans, vmPFC/mOFC has been extensively studied in the context of reward learning (for reviews, see Bartra et al., 2013; Rushworth et al., 2011). In neuro-

imaging studies, expected value signals as computed by a variety of computational models, are consistently observed in vmPFC during value-based decision-making tasks (Bartra et al., 2013). Damage to vmPFC/mOFC in humans and monkeys impairs value-guided decision-making (Camille et al., 2011; Halfmann et al., 2016; Noonan et al., 2010; Rudebeck and Murray, 2014; Rushworth et al., 2011). Some researchers have proposed that vmPFC tracks the value of items regardless of their nature, because vmPFC activation reflects the value across a range of tasks with different reward features from money to aesthetic and social rewards (Behrens et al., 2008; Kim et al., 2011; O'Doherty, 2007; Philiastides et al., 2010). Others have proposed that vmPFC performs value comparisons, because neural signals represent the value difference between alternative options (Boorman et al., 2009; Chau et al., 2014). Value is continually estimated in vmPFC, whereas in other areas of the brain, value computations appear to rely on knowledge or previous experience (Haber and Behrens, 2014; Janowski et al., 2013). This has led to the suggestion that vmPFC may be crucial for selection between alternatives, or at least the transition between valuation and choice selection (Jocham et al., 2012).

The lOFC is another important area for decision-making in the frontal lobe. It is a highly interconnected with both sensorimotor and cognitive control regions (Kringelbach and Rolls, 2004), and could be subdivided into more specific areas. The function of lOFC has been much debated (for a recent review, see Stalnaker et al., 2015), but it is clear that lOFC, like vmPFC/mOFC, has a role in value learning. Whereas the vmPFC/mOFC is active in response to rewards of many different types, cells in lOFC have been shown to differentiate between different types of rewards (Padoa-Schioppa and Assad, 2008). It has been proposed that lOFC is especially important for credit assignment: the process of associating expected values with different visual stimuli during association learning (Jocham et al., 2016; Rushworth et al., 2011; Walton et al., 2010). Additionally, lOFC can represent other aspects detailing a decision-making problem, such as the context, position and reward associated with objects (Farovik et al., 2015).

Recently, the theory emerged that the combined structures of mOFC/lOFC may specifically represent a cognitive map of task-relevant state information (Schuck et al., 2016). This idea integrates the different findings presented by different researchers (Stalnaker et al., 2015), such as OFC encoding stimulus identity related to reward (Howard et al., 2015), but not related to stimulus identity regardless of outcome (Klein-Flügge et al., 2013; Schuck et al., 2016).

The insula has been identified as another important area for decision-making. This is unsurprising, given its central position between cortical areas and the striatum, both topographically and functionally. The insula projects to the NAcc, and receives inputs from sensory as well as prefrontal cortical areas like the vmPFC (Haber and Behrens, 2014). Anticipatory activity in the insula has been

found to signal both anticipation of reward and punishment, and possibly reflects uncertainty (Oldham et al., 2018). Evidence from participants with brain damage demonstrated that after damage to the insula, decision-making is impaired, especially in the aversive domain (Clark et al., 2008, 2014; Von Siebenthal et al., 2017; Weller et al., 2009). Functional neuroimaging studies have found insula activation in the context of weighing losses and risky outcomes (Cox et al., 2008; Rudorf et al., 2012). The insula also likely integrates information with evaluating risk or uncertainty during decision-making (Singer et al., 2009). Additionally, whereas activity in the striatum reliably corresponds to positive RPEs, the insula is robustly active in response to aversive prediction errors (Garrison et al., 2013).

dACC interfaces between reward networks (VS connecting to OFC, insula and vmPFC) and motor networks: dACC functionally couples with the motor cortex responsible for action execution when vmPFC signals the value for that action (Asemi et al., 2015). Therefore, dACC is commonly thought to integrate multisensory evidence. Additionally, recent evidence suggests that ACC plays a role in belief updating, showing a simple preference between one action or another, updating beliefs about how beneficial each action is in real time (Hunt et al., 2018). dACC is also important in overcoming actions costs and exuding the effort needed to achieve a reward (Croxson et al., 2009; Rudebeck et al., 2006). Another area integrated in this network is the dlPFC, which seems to largely play a role in focusing attention on different aspects of the task at hand (Hunt et al., 2018). Lastly, the FPC has been found to play a role in balancing exploration and exploitation (Badre et al., 2012; Raja Beharelle et al., 2015), as discussed in the RL section above. Combined, these cortical areas combine value signals passed from striatum with information about the environment.

Dopamine in the cortex

In the cortex dopamine plays a different role than in the striatum. In PFC, dopamine exerts effects on NMDA and GABA currents in PFC. D1-Rs are more abundant in PFC than D2-Rs (Hall et al., 1994). It has been proposed that D1-R activation augments the robustness of preferred representations in working memory (Seamans and Yang, 2004), although very high activation may reduce robustness, resulting in an inverted U-shape relationship between D1 activation and working memory performance (Cools and D'Esposito, 2011; Sawaguchi and Goldman-Rakic, 1991). Conversely, D2-R activation has been suggested to reduce robustness of working memory representations in the cortex, making them more susceptible to updating with new information (Seamans and Yang, 2004). Dopaminergic modulation of such flexible working memory updating has also been suggested to be dependent on D2-Rs in striatum (Cools and D'Esposito, 2011).

Dopamine, decision-making and aging

As mentioned in the beginning of the introduction, aging has a profound effect on the integrity of the dopaminergic system, as well as the integrity of cortical areas, and the connection between them and subcortical areas (Samanez-Larkin et al., 2012). Thus, aging can be expected to have a widespread effect on value-based decision-making performance as well. There is an abundance of literature that has investigated this.

Value-based learning has been found to be impaired in older adults in multiple studies (Chowdhury et al., 2013a; Eppinger et al., 2011, 2015; Mell et al., 2005; Samanez-Larkin and Knutson, 2015; Samanez-Larkin et al., 2012; Weiler et al., 2008). There are several value-based decision-making tasks one can consider for such observations. One example is reversal learning tasks, where participants have to choose the best option from a set of options. After a while, reward contingencies change, and participants need to reverse their choice contingencies accordingly until contingencies change again. Older adults perform worse on such tasks than younger adults (Mell et al., 2005; Schoenfeld et al., 2014). Additionally, older adults perform worse on probabilistic reward learning tasks, where participants are required to adapt to fluctuating reward payoffs. Within a RL framework, this manifests itself in a slower learning rate (α in the section on RL above) for the updating of stimulus values (Chowdhury et al., 2013a; Rutledge et al., 2009), or reduced sensitivity to punishments.

The anatomy and function of the dopamine system provide an attractive explanation for this deficit. A widely accepted theory is that age-related decline in the dopamine system results in a less pronounced RPE signal (Eppinger et al., 2011; Samanez-Larkin et al., 2014). This would lead to a less efficient teaching signal from the midbrain to the basal ganglia. One fMRI study provided evidence for this theory: older adults showed no expected value (Q) component to their RPE signal, unless their dopamine system was boosted with L-DOPA (Chowdhury et al., 2013a). Additionally, diffusion tensor imaging (DTI) results, which indicate the integrity of the white matter in the brain, showed that the integrity of the connection between SN and striatum accounted for this incomplete RPE (Chowdhury et al., 2013a).

In addition to the decreased ability of the dopamine system to convey RPEs, frontostriatal connections and frontal functioning are both important contributors to decreased value-based decision-making in older adults. Some studies suggested connections between frontal and striatal regions are another main culprit for deficits in probabilistic reward learning in older adults (Samanez-Larkin et al., 2012; Vijver et al., 2016).

How can we study the dopaminergic system?

In order to investigate the role of dopamine in decision-making, we need ways of studying the dopaminergic system in humans. There are several ways to study neurotransmitter systems in humans. Each have their upsides and downsides, and I will briefly lay out why using PET to study the dopaminergic system is of use here.

One possible way of studying neurotransmitter systems is with pharmacological agents that have a known effect on a receptor of a certain type. A great benefit of these types of studies is the possibility to use every individual as their own control. This allows (with some caveats described below) for causal inference about the effect that the manipulation of a neurotransmitter system has. There are many different possible agents that can be used: there are agonists, that activate the receptor type of the target, or antagonists, that inhibit them. Other possible pharmacological agents are precursors of naturally occurring neurotransmitters, of which L-DOPA is an example (Bear et al., 2007).

Although of great value, there are several challenges with the use of receptor agonists and antagonists. As stated before, in order to observe a behavioural change, the same behaviour needs to be measured twice in the same individual. This could lead to retest effect in behaviour, which may be difficult to dissociate from the effect of the drug. Counterbalancing the order of placebo (or drug with an opposing function, e.g. an agonist) and treatment (e.g. an antagonist) can solve part of this problem, but the drug may also modulate the learning process of the behaviour that is investigated. This leads to potential problems with the baseline placebo condition in those participants who receive pharmacological treatment first (Rogers, 2011).

Another challenge is posed by the complicated pharmacology of some receptors targeted by pharmacological studies. Haloperidol is an example of a dopamine D2-R antagonist. If this drug is used to study the dopaminergic system in humans, we know that it will specifically antagonize dopamine D2-Rs (Frank and O'Reilly, 2006; Pessiglione et al., 2006). D2 agonist such as haloperidol has been shown to have a vastly different effects for different dosages (Clatworthy et al., 2009; Huang et al., 2010), which sometimes depend on baseline measures of dopamine functioning (Cools, 2008; Cools et al., 2009). Most pharmacological studies only use one dose and compare the effects of that dose to placebo. Any changes in behaviour as a result of the manipulation can post hoc be explained as a result of these dosage effects. In order to dissociate these effects, researchers are required to also test participants' baseline dopaminergic functioning with additional PET measures, cognitive proxies or genetic analysis (see below), or use a range of different doses.

A challenge that both PET and pharmacological studies face is that radioligands, as well as agonists and antagonists will act on all receptor types available for binding in the brain. For example, D2-Rs exist both on the postsynaptic and presynaptic terminal. Because presynaptic D2-Rs are thought to downregulate dopamine release from the presynaptic terminal, D2-R antagonists could have opposing effects depending on their pre- and postsynaptic workings. This complicates the scientific interpretation of results involving D2-R (ant)agonists (Frank and O'Reilly, 2006; Rogers, 2011; Soares-Cunha et al., 2016). Some studies have suggested that the pre- and postsynaptic workings, as well as the specific effects, may interact with the specific task circumstances, as well as with an individual's cognitive, medication or genetic status (Soares-Cunha et al., 2016).

Lastly, pharmacological challenges are global and it is therefore difficult to investigate specific effects of (for example) striatal and cortical dopamine receptor functioning. Although a single PET scan without drug challenges does not allow for causal manipulations, as pharmacological studies do, the resulting scan has very high spatial resolution. This allows researchers to draw conclusions about the baseline dopamine levels in specific brain areas, and how regionally specific levels of dopamine receptor availability may contribute to decision-making. Ideally, we would use multiple PET scans, genetic tests and pharmacological challenges to obtain a complete picture of dopaminergic functioning, but the resource-heavy nature of such an endeavour precludes such studies.

In addition to (or in combination with) pharmacological manipulations, researchers can investigate neurotransmitter systems by investigating the relationship between behaviour and genetic polymorphisms that predict aspects of that neurotransmitter system. Common examples of such genes for dopamine include catecholamine-O-methyltransferase (COMT) and DRD4 (which are related to levels of prefrontal dopamine), dopamine- and cAMP-regulated phosphoprotein (DARPP-32), DAT1 and DRD2 (which are related to levels of striatal dopamine and striatal dopamine receptor density). Such studies have provided a great amount of insight in the interactions between genotypes and behaviour on a range of tasks (Frank and Fossella, 2011).

The main challenge for such genetic investigations lies in the small effect that a single genetic polymorphism has in steering neural activity and behaviour. This implies that sample sizes in such studies need to be large, and the discovery of new genes that affect behaviour should be subject to careful scrutiny in relation to previous literature and the presumed mechanism at work. In addition, genetic effects interact with the environment and with other genes, which can complicate the investigation of such effects, or allow for great post-hoc rationalization of any observed effects. Nonetheless, effects of all of the genetic polymorphisms mentioned above have been replicated across tasks and studies, demonstrating the value of this approach (for a review, see Frank and Fossella (2011)).

We used PET in the current set of studies, because we were interested in disentangling the effects of cortical and striatal dopamine receptor availability. In addition, we wanted to investigate the age-related effects of dopamine receptor deterioration, into which genetic factors do not provide an insight. However, it should be noted that all three approaches (pharmacological manipulations, genetic investigations and PET studies) provide insights into the workings of the human dopaminergic system, and that all of them (ideally in combination with each other) should be used to expand our knowledge about it.

AIM OF THE THESIS

The aim of this thesis was to investigate the role of endogenous dopamine D1-R availability, and its age-related deterioration, in neural and behavioural mechanisms of decision making. I present four specific sub-goals:

1. To study how aging affects neural processing (as measured with fMRI/BOLD) of value anticipation and RPEs in striatum and PFC and how these effects are related to the dopamine D1-R availability in cortical and subcortical regions.
2. To study how aging affects the structural connectivity between NAcc and PFC, how this structural connectivity is related to the expression of expected value in the cortex, and how the two affect value-based decision-making.
3. To study how aging and dopamine D1-R availability are related to motivational biases during learning, and whether these biases are related to the dopamine D1-R availability in either cortical or subcortical regions, or both.
4. To study whether different anticipatory and outcome processing patterns of brain activity (as measures with fMRI/BOLD) during a valenced Go/NoGo are supported by dopamine D1-R availability in cortical and subcortical regions, and how aging affects these outcome processing patterns.

METHODS

Research participants and procedure

The studies that are a part of this thesis all use data from the same 60 participants in the dopamine and decision-making (DAD) sample. These include 30 older (age 65-75 years, 12 females, mean age = 70.7 years, standard deviation (SD) = 2.75 years) and 30 younger (age 19-32 years, 18 females, mean age = 24.2 years, SD = 3.44 years) participants. The research was performed at the Umeå Functional Brain Imaging Centre, and the Radiology Department at Umeå University. Ethical permission was obtained from the Umeå Ethical review board. All participants were right-handed, and had normal or corrected-to-normal vision. Before recruitment into the study sample, candidate participants were screened by research nurses with the use of a screening questionnaire. Candidates were excluded if they had a history of substance abuse, mental illness, diabetes mellitus, or used more than 2 medications for arterial hypertension. All participants provided written informed consent prior to commencing the study. Participants were paid 2000sek (~$225) for participation in the study, and had the opportunity to earn additional money in each of the decision-making tasks (1 sek per obtained reward).

Participants came in on two different days, with 3-44 days between them. On the first day, they performed two value-based decision-making tasks in a magnetic resonance imaging (MRI) scanner, and three decision-making tasks outside the scanner, including all the tasks presented in this thesis. On the second day, participants performed three more decision-making tasks. On this day, they also underwent a PET/CT scan with the radiotracer [^{11}C]SCH23390, a ligand that binds to dopamine D1-Rs. PET imaging failed for one older participant, due to the radiotracer not entering the bloodstream. In addition, one older participant felt unwell during the PET imaging and withdrew from that part of the study. Those participants' fMRI and behavioural data are still included where possible.

Study 3 includes data from two additional datasets. First, a dataset that was previously published in Chowdhury et al., 2013b. This dataset included 42 older participants (age 64-75, 29 females, mean age = 69.1 years, SD = 3.44 years) and 47 younger participants (28 females, mean age = 23.1 years, SD = 4.1 years). We used behavioural data on the learning version of the Go/No-Go task to validate the computational modelling analysis used in our own dataset. Second, a PET dataset previously published in Rieckmann et al., (2011). This dataset included 20 older participants (10 females, mean age = 70.4 years, SD = 3.12 years) and 20 younger participants (10 females, mean age = 25.2 years, SD = 2.21 years), for whom ROI data was available for a range of regions, imaged with [^{11}C]SCH23390. We used dopamine D1-R availability data from this dataset to validate the principal component analysis performed in the DAD dataset used in **study 3** and **4**.

Task descriptions

This thesis includes studies that use performance on two different value-based decision-making tasks as an outcome variable: a two-armed bandit (TAB), and a valenced Go/No-Go task. I will briefly describe each task below.

The **two-armed bandit task** (figure 4) is a commonly-used task in decision-making research. Participants performed this task in the MRI scanner. Stimuli were presented on a computer screen that participants viewed through an angled mirror above their heads. At the start of each trial, participants were first presented with a fixation cross in the centre of the screen, and then two fractal images, which each representing one bandit "arm". They could select one of the two arms through a button press. After the arm was selected, a rectangle would frame the chosen option, and after a variable interval (3000ms-RT), the outcome was presented on the screen. When the outcome was a reward, participants saw a green arrow pointing upwards. When the outcome was a reward omission, participants saw a yellow horizontal bar. If participants pressed the button after the response window had ended (3000ms), they saw a red X on the screen, and the text "you were too slow!" in Swedish. The next trial then started after the next inter-trial-interval. The outcome probability of receiving a reward for each arm varied over the course of the experiment according to a random Gaussian walk (figure 4).

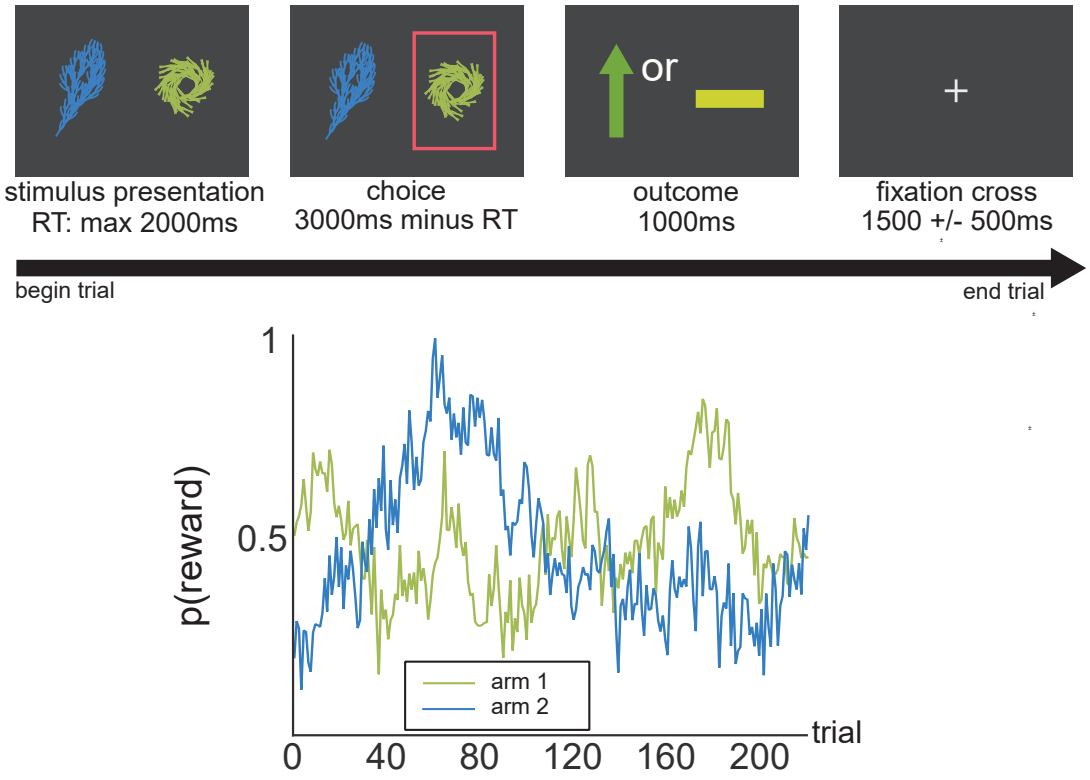

Figure 4. *Top panel: schematic of the TAB and timing. Bottom panel: random Gaussian walks used in this TAB.*

The probability of receiving a reward from bandit arm *a* on trial t was between 1 and 100 percentage points, drawn from a Gaussian distribution (standard deviation $\sigma = 4$) around a mean $\mu_{a,t}$ rounded to the nearest integer. We calculated the Gaussian walk identically to Daw et al., 2006: The starting value of the Gaussian random walk was equal to $\mu_{a,1}$, which was randomly drawn from a uniform distribution between 1 and 100. At each timestep *t*, the means μ for each bandit arm diffused in a decaying Gaussian random walk from this initial point, with

$$\mu_{a,t+1} = \lambda\mu_{a,t} + (1-\lambda)\theta + \nu \quad (3)$$

for each bandit arm *a*. The decay parameter λ was 0.9836, the decay center θ was 50, and the diffusion noise ν was zero-mean Gaussian (standard deviation $\sigma_d = 2.8$) (Daw et al., 2006). Percentage points were divided by 100 to obtain probability values (figure 4).

Participants were instructed to try and maximise the number of rewards during the duration of the task, and were aware that the probability of reward receipt could vary for each fractal image. Figure 4 summarizes the trial timeline for the experiment, and shows the random Gaussian walks used in **study 1** and **2**. The same Gaussian walks were used for each participant, but the assignment of Gaussian walk to bandit arm identity was counterbalanced between participants. In total, participants performed 220 trials, with a self-paced break after 110 trials.

The **Go/No-Go task** consists of four conditions. The four conditions were "go to win" (GW), "go to avoid losing" (GAL), "no-go to win" (NGW), and "no-go to avoid losing" (NGAL). Each trial started with the presentation of a single fractal image on the screen. Four possible different fractals could be presented, one for each condition. After the fractal, a target was presented, which could be on the left or right side, relative to the fixation cross. Participants then had to decide between performing an active choice, in this case a "go", which entailed pressing a button on the side that they had seen the target, or an inactive choice, or a "no-go", which entailed not pressing a button at all. Participants were explicitly instructed not to press on the side that the target was not presented on. If they did, responses were still counted as a "go".

The Go/No-Go task (figure 5) consisted of two phases, a **learning phase (study 3)** and an **instructed phase (study 4)**. The learning phase was performed outside the scanner, where participants were seated in front of a laptop computer in a quiet room. Participants saw 45 trials of each condition during this phase. They were unaware of the correct choice-outcome contingencies, and were instructed that active or passive choices could be correct for each image. A fixed time after the choice was made, the outcome was presented on the screen. This would either be "win", represented as a green arrow pointing upwards, "nothing", represented

by a yellow horizontal bar, or "lose", represented by a red arrow, pointing downwards. In the GW condition, active choices would lead to a win 80% of the time, and to nothing 20% of the time (and vice versa for inactive choices). In GAL, active choices would lead to nothing 80% of the time, and to a loss 20% of the time. Conversely, in NGW, inactive choices would lead to a win 80% of the time, and to nothing 20% of the time (and vice versa for active choices). In NGAL, inactive choices would lead to nothing 80% of the time, and a punishment 20% of the time. Fractal identities and conditions were randomized between participants. Figure 5 represents the experiment timeline, the different conditions and their outcome contingencies. They were instructed to find out through trial and error which choice was optimal for each image. Additionally, participants were aware that the outcome contingencies were probabilistic.

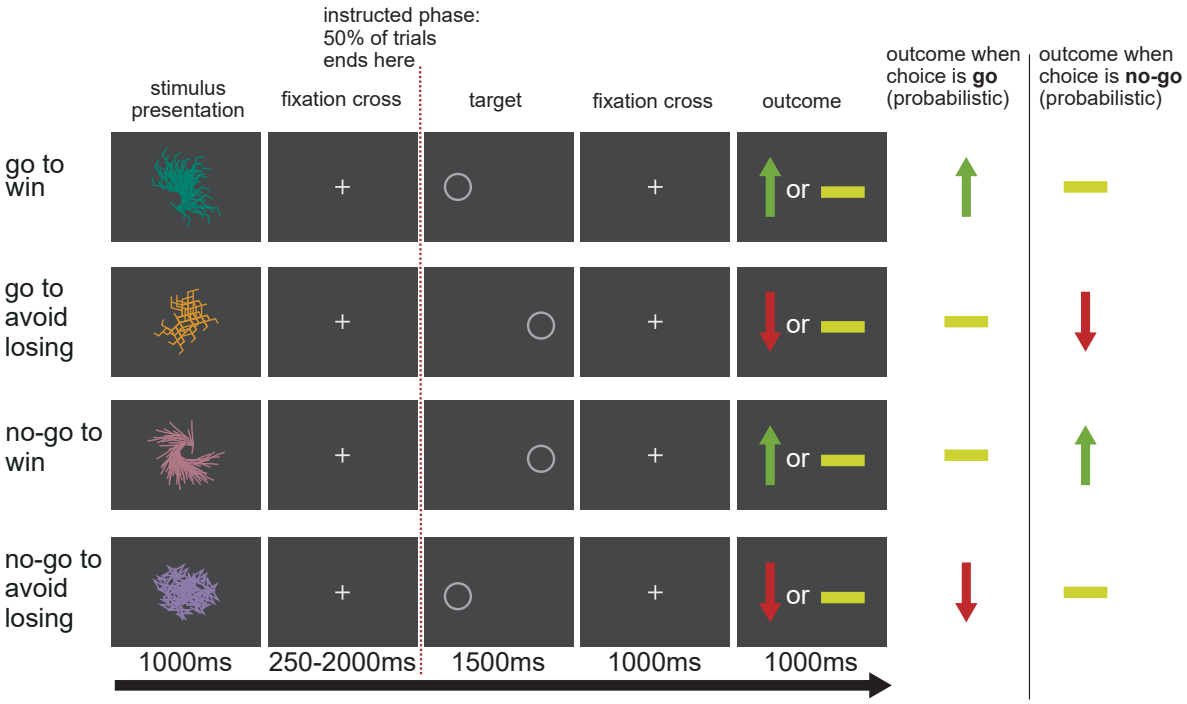

Figure 5. *The four conditions of the go/no-go task and the choice/outcome contingencies. Targets were presented on the left or right side of the screen, and the side randomly varied over the course of the experiment. In the learning phase, feedback was presented probabilistically with 80% reliability, in the instructed phase this was lowered to 70%. 50% of the trials in the instructed phase ended after stimulus presentation.*

After the learning phase, participants performed the instructed phase. Note that during the instructed phase, the probabilities of contingent feedback were changed to 70/30 instead of 80/20. Before starting the instructed task, participants were explicitly instructed about the choice-outcome contingencies for each fractal, with a message on the screen displaying what the correct choice

was for each stimulus. After this instruction, participants performed 10 trials in each condition, where after the outcome was presented on the screen, they also saw text, explaining what the correct choice was for the preceding trial. After this, participants performed a block of 10 trials in each condition where half of the trials ended after stimulus presentation, and written feedback was omitted. The goal of this set of instructions was twofold: first, it was designed to ensure perfect performance in the next phase of the task, the instructed phase. Second, it familiarised participants with the fact that during the instructed phase, some trials would end before target detection.

The instructed phase was performed inside the MRI scanner. Participants saw a total of 60 trials per condition during the instructed phase. Here, the same assignment of fractals to action-outcome contingencies was used as during the learning phase. 30 of the 60 trials were cut off after the presentation of the fractal (Figure 5, so-called "sham" trials), meaning participants never saw the target or the outcome. This was done to dissociate the action execution component from the anticipatory component during fMRI analysis. The total of 240 trials were randomized and presented in three blocks of 80 trials, with self-paced breaks in between.

Computational modelling

The computational models are constructed to approximate the value-estimation process that is presumed to occur in the human brain. The action propensities ($m_a(t)$) for each choice $a \in \{0,1\}$ on trial t in the two tasks were calculated according to one of the two learning model families, Bayesian and RL. These choice values were then entered into the softmax equation, which calculated the probability that choice c on trial t would be action a:

$$P(c_t = a) = \frac{exp[\beta m_a(t)]}{\sum_{b=1}^{n} exp[\beta m_b(t)]} \quad (4)$$

Below I will briefly describe the most important models and the corresponding learning model for each task, and each considered dataset. However, in the analysis procedure, we investigated the different combinations of parameters these parameters to obtain a fit.

Two-armed bandit – Rescorla-Wagner (study 1)

We used two different families of computational models to model the TAB task data: Rescorla-Wagner models and Bayesian observer models. Although a Bayesian model outperformed the Rescorla-Wagner model in **study 1** and **2**, both families of learning model are useful for calculating action propensities on this task. Within the Rescorla-Wagner model family, the best model included the

Rescorla-Wagner updating rule for the chosen option, akin to the learning rule presented in the introduction:

$$Q_{t+1}(a_t) = Q_t(a_t) + \alpha \cdot \delta_t \quad (5)$$

Where α was the learning rate, δ_t was the prediction error on that trial, and $a \in \{0,1\}$, reflecting each possible action, or each bandit arm. We included an additional parameter in the definition of $m_a(t)$: a perseveration parameter b (with $-\infty < b < \infty$). This parameter raises or lowers the expected value of a stimulus if that stimulus was also chosen on the previous trial. Thus,

$$m_a(t) = Q_a(t) + b\chi_{a=a(t-1)} \quad (6)$$

where a positive value of b reflects a tendency to perseverate (repeat the same choice), and a negative value reflects avoiding perseveration. Lastly, a forgetting rate parameter φ was added to the updating rule for the unchosen option:

$$Q_{t+1}(1 - a_t) = Q_t(1 - a_t) + \varphi(0.5 - Q(1 - a_t)) \quad (7)$$

Which determined the speed with which the value of a repeatedly unchosen bandit arm would be relaxed towards 0.5.

Two-armed bandit – Bayesian observer model (study 1 and 2)

In the Bayesian observer model, the probability of obtaining a reward for each possible action $a \in \{0,1\}$ (corresponding to each bandit arm) was represented as a beta distribution:

$$\theta_a \sim \beta(\theta_a; \gamma_a, \varepsilon_a) \quad (8)$$

The outcome on each trial lead to the updating of each distribution. From these distributions, we can mathematically derive the mean probability of obtaining a reward (which we refer to as $Q_a(t)$, for consistency with the RL models) and the variance in reward probability ($V_a(t)$):

$$Q_a(t) = \frac{\gamma_a}{(\gamma_a + \varepsilon_a)} \quad (9)$$

$$V_a(t) = \frac{\gamma_a \varepsilon_a}{(\gamma_a + \varepsilon_a)^2(\gamma_a + \varepsilon_a + 1)} \quad (10)$$

The starting values of the beta distribution parameters were set to 1 ($\gamma_a = \varepsilon_a = 1$). This implies that at the beginning of the experiment, $Q_0(1) = Q_1(1) = 0.5$ and maximum variance $V_0(1) = V_1(1) = 0.083$ reflecting a chance expectation of reward for both bandit arms and a maximum uncertainty about the underlying probability

distributions. If action a lead to a reward, γ_a was increased by 1, and γ_{1-a}, ε_{1-a}, and ε_a are relaxed towards 1. After a neutral outcome (reward omission), ε_a was increased by 1 and γ_a, γ_{1-a} and ε_{1-a} are relaxed towards 1:

$$\gamma_{a(t)}(t+1) = (1-\omega)\gamma_{a(t)}(t) + \omega + 1; \quad \text{and}$$
$$\varepsilon_{a(t)}(t+1) = (1-\omega)\varepsilon_{a(t)}(t) + \omega; \quad if\ R(t) = 1 \tag{11}$$

$$\gamma_{a(t)}(t+1) = (1-\omega)\gamma_{a(t)}(t) + \omega; \quad \text{and}$$
$$\varepsilon_{a(t)}(t+1) = (1-\omega)\varepsilon_{a(t)}(t) + \omega + 1; \quad if\ R(t) = 0 \tag{12}$$

For the unchosen bandit arm:

$$\gamma_{1-a(t)}(t+1) = (1-\lambda)\gamma_{1-a(t)}(t) + \lambda; \quad \text{and}$$
$$\varepsilon_{1-a(t)}(t+1) = (1-\lambda)\varepsilon_{1-a(t)}(t) + \lambda; \tag{13}$$

ω and λ are individual participants' free parameters that determine the speed with which reward probabilities are updated (ω, with $0 < \omega < 1$) and forgotten (λ, with $0 < \lambda < 1$).

Beyond ω and λ, which determined the value of $Q_a(t)$, two additional parameters were added to estimate the action propensity $m_a(t)$ for the two bandit arms. First, the variance of the bandit arm that was not chosen on trail t was added to the action propensity of that bandit arm on trial t + 1:

$$m_a(t+1)\chi_{a=1-a(t)} = Q_a(t+1)\chi_{a=1-a(t)} + \upsilon V_a(t+1)_{\chi_{a=1-a(t)}} \tag{14}$$

If $\upsilon > 0$, choices with high variance are favoured and when $\upsilon < 0$, choices with high variance are avoided. Lastly, a measure of confidence was added to the value of the bandit arm that was not chosen on trial t. Relative confidence was defined as the probability that a sample drawn from the distribution for bandit arm a would be more likely to lead to a reward than a sample drawn from the distribution for bandit arm $1-a$. This relative confidence, calculated at trial t, was added to the unchosen option at trial t + 1:

$$m_a(t+1)\chi_{a=1-a(t)} = Q_a(t+1)\chi_{a=1-a(t)} +$$
$$\upsilon V_a(t+1)_{\chi_{a=1-a(t)}} +$$
$$\kappa C_{rel}(t) \tag{15}$$

Where κ was an individually fitted parameter that weighted the relative confidence C_{rel}. Relative confidence was calculated by calculating the probability that a random sample drawn from the beta distribution of bandit a would be more likely to lead to a reward than a sample drawn from the beta distribution of bandit *1-a*.

Given that our Bayesian observer model tracks subjective estimates of the full probability distribution of obtaining a reward for each bandit arm, the relative probability of one bandit arm being better than another can be approximated by:

$$C_1(t) = P(\theta_1 > \theta_0) = \int_{\theta_1=0}^{1} d\theta_1 \beta(\theta_1;\, \gamma_1, \varepsilon_1) \int_{\theta_0=0}^{\theta_1} d\theta_0 \beta(\theta_0;\, \gamma_0, \varepsilon_0) \quad (16)$$

$$C_0(t) = P(\theta_0 > \theta_1) = 1 - C_1(t) \quad (17)$$

Given the simple relationship between these two confidences, there are various essentially equivalent ways of incorporating it into choice. We considered the relative confidence in the choice on a trial:

$$C_{rel}(t) = P\left(\theta_{a(t)} > \theta_{1-a(t)}\right) - P\left(\theta_{1-a(t)} > \theta_{a(t)}\right) = 2P\left(\theta_{a(t)} > \theta_{1-a(t)}\right) - 1 \quad (18)$$

and assessed the extent to which the relative confidence on trial *t*–1 encouraged switching on trial *t* by adding a factor $\kappa C^{\mathrm{rel}}(t-1)\chi_{a=1-a(t-1)}$ to the action that was not chosen on trial *t*–1. Here, positive values of κ make the subjects more likely to switch if they had been more confident – i.e., reflecting a tendency to believe that the 'grass might be greener on the other side'.

In the TAB, the Bayesian observer model outperformed the Rescorla-Wagner model.

Go/No-Go – Rescorla-Wagner (study 3)

For the Go/No-Go task, we compared a range of models of the Rescorla-Wagner family only. Value computations for this task were calculated differently, because participants had the choice between performing a Go or a NoGo in response to each of four stimuli they were presented with, instead of a choice between stimuli. For that reason, we modelled the value of a "go" choice and a "no-go" choice separately for each stimulus. The result of the softmax computation would then be the weighted relative probability of performing a go, versus performing a no-go. In addition to including the action weight as in formula 1, we added an irreducible noise parameter ξ, which made the softmax rule robust to "flukes", where participants would show a sudden diversion from action propensities. This parameter can also be referred to as a "lapse rate". The altered softmax then reads:

$$P(a_t, s_t) = \left[\frac{exp[W(a_t,s_t)]}{\sum_{a'} exp[W(a'.s_t)]}\right](1-\xi) + \frac{\xi}{2} \quad (19)$$

The action weight $W(a_t,s_t)$, which was tracked for each action $a \in \{0,1\}$, with 1 for "go" and 0 for "no-go", and each state $s \in \{1,2,3,4\}$, reflecting the stimulus identity on that trial *t*.

In order to keep the notation in this modelling section consistent with the publication (**study 3**), and because the learning rate and softmax temperature parameters

are subject to modification by model parameters in the modelling of this task, I will use different symbols for these in this model compared to the TAB model. However, theoretically ε and ρ perform identical roles in this learning rule to α and β in the modelling of the TAB. All models we considered included the value Q of each action as determined by the Rescorla-Wagner updating rule:

$$Q_t(a_t, s_t) = Q_{t-1}(a_t, s_t) + \varepsilon\big(\rho r\ - Q_{t-1}(a_t, s_t)\big) \tag{20}$$

All models included a learning rate ε. Rewards, neutral outcomes and punishments were entered in the model as $r \in \{-1, 0, 1\}$. ρ reflected weighting of reward and punishment, determining the effective size of the reward or punishment. In some models, ρ took on separate values for rewards and punishments, assuming that forgoing a reward could be more or less aversive than obtaining a punishment, a model definition that has been shown to consistently improve model fit (Cavanagh et al., 2013; Guitart-Masip et al., 2014a). A single value for ρ, symmetrically weighting rewards and punishments, would have the same effect as a single temperature parameter β as in equation 4.

The effect we intend to capture modelling the go/no-go task is the slowed learning for passive choices that lead to rewards, and relative slower learning of active choices leading to avoiding punishments (compared to active choices leading to rewards and passive choices to avoid punishments). In order to do this, we add three additional parameters to the models in different constellations.

First, go/no-go models included an individually fitted static bias parameter b that was added to the value of the action "go", regardless of the potential reward of that trial type. Thus, for every trial type:

$$W_t(a_t, s_t) = \begin{cases} Q_t(a_t, s_t) + b & if\ a = go, else \\ Q_t(a_t, s_t) \end{cases} \tag{21}$$

Second, some go/no-go models included a Pavlovian term. This term added the expected value on the current state ($V_t(s_t)$) to the value of go choices. The term was weighted by an individually fitted Pavlovian parameter π:

$$W_t(a_t, s_t) = \begin{cases} Q_t(a_t, s_t) + b + \pi V_t(s) & if\ a = go, else \\ Q_t(a_t, s_t) \end{cases} \tag{22}$$

where $\pi \geq 0$.

For models that included a Pavlovian factor, V was computed as

$$V_t(s_t) = V_{t-1}(s_t) + \varepsilon\big(\rho r_t - V_{t-1}(s_t)\big) \tag{23}$$

The Pavlovian parameter devalued the value of go choices in punishment conditions, proportionally to the value of the stimulus (V(s)), which was negative in these instances. For rewards, the Pavlovian parameter boosted the value of go choices in proportion to the positive value of the stimuli that signalled potential rewards.

Finally, we added an instrumental learning bias κ to some models. κ modulated the participants' learning rate ε depending on whether the choice was active or passive, and subsequent feedback was positive or negative. In its most elaborate form, the value of this parameter was added to ε on trials that resulted in a rewarded go response, and subtracted from ε on trials that resulted in punished no-go choices.

$$\varepsilon = \begin{cases} \varepsilon_{RewardedGo} = \varepsilon_0 + \kappa \\ \varepsilon_{PunishedNoGo} = \varepsilon_0 - \kappa \\ \varepsilon_{other} = \varepsilon_0 \end{cases} \quad (24)$$

The winning model for the main sample in **study 3** included ρ_{win}, ρ_{lose}, ε, ξ and $\kappa_{\text{rewarded-go}}$. In this model, κ modulated ε only on rewarded go trials. Additionally, we reported the winning model in a dataset from a previously published dataset (Chowdhury et al., 2013b). In that dataset, the winning model included ρ_{win}, ρ_{lose}, ε, ξ, π and κ on rewarded go trials as well as punished no-go trials.

Model fitting

Model parameters for all tasks were fitted using the statistical procedure known as expectation-maximization. We used MATLAB's *fminsearch* function to perform Laplacian approximation of a maximum a posteriori estimates for all parameters for all participants. The expectation step defined the probability distribution around a vector of means $\boldsymbol{h}$ and variances Σ:

$$p(h|A_i) \approx \mathcal{N}(h_i^{(k)}, \Sigma_i^{(k)}) \quad (25)$$

$A = \{A_i\}_{i=1}^{N}$ represents all the actions by all the N subjects. N denotes a normal distribution and $\Sigma_i^{(k)}$ is the second moment around $h_i^{(k)}$. h is a vector of parameter estimates (means) for participant i at step k. $\Sigma_i^{(k)}$ approximates the variance around the parameter estimates. Maximization then occurred in k steps, where

$$h_i^{(k)} = argmax_h \, p(A_i|h)p(h|\theta^{k-1}) \quad (26)$$

Where at the starting point, $\theta = \{m, v^2\}$ were the prior mean (0) and variance (2^2), which served a regularizing purpose for parameters, preventing them from taking on extreme values. Where appropriate, parameter values were transformed into model space with log and inverse sigmoid transforms: parameters constrained to be between 0 and 1 were transformed with an inverse sigmoid transform, and parameters constrained to be positive were constrained with a logarithmic transform. Updates for the mean and variance of the group level parameters on step k occurred as follows:

$$m^{(k)} = \frac{1}{N}\sum_i h_i^{(k)} \quad (27)$$

$$(v^{(k)})^2 = \frac{1}{N}\left[(h_i^{(k)})^2 + \Sigma_i^{(k)}\right] - (m^{(k)})^2 \quad (28)$$

Parameter values were considered for each participant's decisions, until the difference between considered values stayed below 0.01 between steps, which was the requirement for model convergence. As stated above, in the first instance, all parameters started with Gaussian priors with mean = 0 and SD = 2. After that, the maximum a posteriori estimates of the group mean and variance were used as a prior for the next instance. In order to avoid local minima, we performed 10 instances of this model fitting procedure to obtain maximum a posteriori estimates, and selected the instance where the highest total likelihood for all subjects was obtained.

Model selection

BICs were calculated for the instance where the maximum likelihood for each model was observed. As we believe all considered models to be equally likely a priori, we examined the model log likelihood $log\, p(M|A)$ for each model M given all the data A directly, without considering a prior for model selection. The model log likelihood could be approximated in two steps. First, we approximated the integral over the hyperparameters at the group level, using the Baysian Information Criterion:

$$logp(A|M) = \int d\theta p(A|\theta)p(\theta|M) \approx -\frac{1}{2}BIC_{int} =$$
$$logp\left(A\middle|\hat{\theta}^{ML}\right) - \frac{1}{2}|M|\log(|A|) \quad (29)$$

where |A| is the total number of choices made by all subjects, and |M| is the number of parameters fitted, contributing to the penalty parameter for the iBIC. $logp\left(A\middle|\hat{\theta}^{ML}\right)$ is not the sum of individual likelihoods, but the sum of the integrals over the individual parameters. Therefore, we refer to this value as the BIC_{int}, or iBIC, with i for "integral". This integral was approximated by sampling from the fitted parameter values:

$$logp\left(A\middle|\hat{\theta}^{ML}\right) = \sum_i \log \int dh\, p(A_i|h)p\left(h|\hat{\theta}^{ML}\right) \approx \sum_i \log \frac{i}{K}\sum_{k=i}^{K} \mathrm{p}(A_i|h^k) \quad (30)$$

where K was set to 2000, and h^k were parameter values drawn independently from the estimated maximum a priori estimate and variance of the group level hyperparameters.

PET imaging

Positron emission tomography (PET) is a neuroimaging technique that allows for the reconstruction and quantification of 3-D images that display neurochemical properties. During a PET imaging session, a radioactive tracer is injected into the bloodstream of the research participant. The radiotracer is a medical compound, of

which an atom has been replaced with a decaying isotope, in our case carbon-11 ([^{11}C]). The medical compound will bind to the target it has been designed to bind to, in our studies dopamine D1-Rs. The isotope that is part of the tracer decays. Isotopic decay results in the release of positrons. These positrons usually travel a short distance in the tissue ($<$1mm), before encountering an electron. When a positron and an electron collide, both particles are annihilated, a reaction which causes two photons to be released in approximately opposite directions (180°) from each other. The PET scanner, which surrounds the area that is imaged, contains detectors that register these opposite photon emissions. From these registrations, a 3-D image is reconstructed. These different reactions across regions provide data on the total concentration of radioactivity in a certain area or interest. A time-activity curve (TAC) shows the average concentration of radioactivity in all voxels of a region of interest over several timeframes.

PET image acquisition

PET images were acquired on a 690 PET/CT scanner (GE Medical Systems, WI, US). A low-dose helical CT scan (20 mA, 120 kV, 0.8s/revolution) was used for PET attenuation correction. Individually fitted thermoplastic masks were used to fixate the participants' heads (Positocasts Thermoplastic; CIVCO medical solutions, IA, US). At the time of an intravenous bolus injection of 200MBq of [^{11}C] SCH23390, a 55 minute dynamic acquisition started (9x120s, 3x180s, 3x260s and 3 x 300s), totaling 18 frames. Attenuation- and decay-corrected 256x256 pixel transaxial PET images were reconstructed to a 25cm field-of-view using the Sharp IR algorithm (6 iterations, 25 subsets, 3.0mm Gaussian post filter). Sharp IR is an advanced version of the Ordered Subset Expectation Maximization (OSEM) method for improving spatial resolution (Ross and Stearns, 2010). The Full-Width Half-Maximum (FWHM) resolution was 3.2mm. This protocol resulted in 47 tomographic slices per timeframe, with 0.98x0.98x3.3mm^3 voxels. Images were decay-corrected to the start of the scan.

PET analysis

We used a ROI-based protocol to estimate non-displaceable binding (BP_{ND}). BP_{ND} values were obtained by coregistering the PET time series images to the T1-weighted MRI images using SPM. From the T1-weighted images, we segmented ROIs using the FIRST algorithm as implemented by FSL (Patenaude et al. 2011). The cerebellum was segmented with the use of Freesurfer's *recon-all* algorithm (Desikan et al. 2006) and used as a reference tissue due to the lack of dopamine D1-Rs in this structure (Hall et al. 1994). The average time activity curves (TAC) were extracted across all voxels within each ROI. Then, BP_{ND} was calculated with the use of the Logan method (Logan et al. 1996) as implemented in imlook4d

(imlook4d version 3.5, https://sites.google.com/site/imlook4d). BP_{ND} values were averaged across hemispheres for the NAcc. BP_{ND} values were taken as a proxy of dopamine D1-R availability.

Definition of regions of interest for PET

Regions of interest were selected based on their relevance to decision-making in previous literature. In **study 1**, we used cortical regions of interest from the freesurfer recon-all parcellation pipeline. After visual inspection of the subcortical parcellation using recon-all, we performed an additional subcortical parcellation with the FIRST algorithm implemented in FSL. This parcellation provided a better fit to subcortical structures. Specific investigations of ROIs are discussed in the individual studies.

Principal Component Analysis

In **study 3** and **4**, we used a principal component analysis (PCA) to tease apart variance in dopamine D1-R signal from the PET data. Because BP_{ND} values in all the ROIs considered were highly collinear (correlation coefficients $r > 0.45$, $p < 0.001$), it is difficult to assess specificity when multiple significant correlations between BP_{ND} values in ROIs and behaviour are observed. In order to obtain orthogonal and meaningful components, we performed a PCA, followed by a varimax rotation. Before performing the PCA, we age-corrected the BP_{ND} values, as age provides a large portion of BP_{ND} variability, and we were interested in capturing potential anatomical, functional or topographical patterns of organization with our PCA (Haber and Knutson, 2010). BP_{ND} values were age-corrected by calculating the effect of age on the BP_{ND} and correcting for this effect in each ROI (Raz et al., 2004):

$$BP_{ND}\text{-adj(participant)} = BP_{ND}(\text{participant}) + \beta_{age}*\text{age(participant)} \quad (31)$$

The number of components to retain was determined with the use of a Catell-Nelson-Gorsuch test (Cng) on the eigenvalues, done with the R package *nFactors* (function nCng, (Gorsuch, 2014)). Cng involves computing the slopes between the eigenvalues in the scree plot. The point at which the greatest change in slope is observed is the cut-off point for the number of components.

Magnetic resonance imaging (MRI)

Magnetic resonance imaging (MRI) is a non-invasive imaging technique that uses a magnetic field to construct 3d anatomical images. Creating an MRI image relies on the different magnetic properties of these anatomical structures. The scanner creates a strong magnetic field, within which a transient varying magnetic fields

are introduced using a radiofrequency pulse. This aligns the nuclei of the atoms inside the scanner into the same orientation. When this transient magnetic field is removed, the atoms in different anatomical structures take a different amount of time to relax into their original state (the spin-lattice relaxation time). In the brain, this measure is used to visualize the difference between white and grey matter. This contrast is also called the T1 contrast, and has been used in all of our studies for the segmentation of the brain into different ROIs for all participants.

Diffusion-weighted imaging (DWI)

Diffusion-weighted imaging is a type of MRI. The goal of DWI is to measure the directionality of water molecules' movement (also called anisotropic diffusion) in the brain's white matter microstructure. The assumption is that stronger anisotropic diffusion of the water molecules indicates higher structural integrity of the white matter microstructure, such as the myelin and the glial cells surrounding the neurons' axons. Instead of a single radiofrequency pulse, during a DWI sequence, radiofrequency pulses are applied in a number of directions (32 in **study 2** in this thesis). These radiofrequency pulses are then relaxed, and the shift of the water molecules between the radiofrequency pulses is measured. This shift can be represented as a 3d ellipsoid called a tensor, which has a Cartesian x, y and z direction in each voxel, of which the three main directions are represented as three eigenvalues. As the main outcome measure of white matter microstructure integrity we use Fractional Anisotropy (FA), defined as the intravoxel preferred directionality of water molecule translational random motion. This is expressed as a ratio ranging from 0 to 1, with 0 being isotropic (non-directional) and 1 being unidirectional. FA is the normalized mean of the three eigenvalues of the tensor as obtained from the DWI sequence:

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{\left(\lambda_1 - \hat{\lambda}\right)^2 + \left(\lambda_2 - \hat{\lambda}\right)^2 + \left(\lambda_3 - \hat{\lambda}\right)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{32}$$

Where $\hat{\lambda}$ is the average of the three eigenvalues, and λ_1, λ_2, and λ_3 are the three eigenvalues (Hagmann et al., 2006).

DWI preprocessing and analysis

Diffusion weighted scans were corrected for motion and current-induced distortions with FSL's eddy_correct. To further correct for geometric distortions the images were non-linearly aligned with the T1-weighted structural scan (Wu et al., 2008) with the ANTs software ("ANTs by Stnava" 2019). Tractograms were generated with the MRtrix software (Tournier et al., 2012) and filtered with the SIFT2 method (Smith et al., 2015), using anatomically-constrained tractography

(Smith et al., 2012). We specified the two inclusion regions of interest (vmPFC and NAcc) as binary mask images, and accepted only streamlines that traversed both inclusion regions. We sampled until we recovered 100 streamlines between the vmPFC and NAcc regions of interest. Subjects were excluded if >200 million streamlines were considered, but less than 20 were selected as probable (n = 11). FA maps were calculated using FSL's dtifit. The tract formed by the reconstructed streamlines was used to mask the FA image, and the average within the tract became the individual's measure of accumbens-to-frontal white matter integrity.

fMRI

When we try to understand the brain, we would like to know which neurons are active during a range of cognitive processes. The ultimate measure of neuronal activity is to record it directly from firing neurons inside the living brain. Unfortunately, techniques that allow for the recording of activity directly from cells (single-cell recordings) are hugely invasive and almost never used in humans (although becoming more common with the emergence of intracranial electroencephalography (EEG) (Shokoueinejad et al., 2019)). With functional MRI, a proxy for neuronal activity is measured with the Blood-Oxygen-Level-Dependent (BOLD) contrast (Ogawa et al., 1990). This is a widely used technique in cognitive neuroscience. The BOLD contrast makes use of oxygen consumption as a proxy for energy expenditure in the brain. As cells consume oxygen, the relative concentration of oxygenated and deoxygenated blood in the blood flow towards those cells changes. Because oxyhemoglobin and deoxyhemoglobin have different magnetic properties, this change in relative concentration can be measured with functional MRI (fMRI). We applied fMRI and the BOLD contrast in **study 1 and 4**.

fMRI acquisition

Brain images were acquired on a MR750 3T scanner (GE Medical Systems, WI, US), equipped with a 32-channel phased-array head coil. T1-weighted 3D-SPGR images were acquired using a single-echo sequence (voxel size: 0.5 x 0.5 x 1 mm, TE = 3.20, flip angle = 12 deg). Diffusion weighted imaging scans were acquired with a spin-EPI T2-weighted sequence (64 slices, voxel size = 1 x 1 x 2 mm, TR = 8000 ms, TE = 84.4ms, FoV = 25 cm, flip angle = 90°), using 3 repetitions, with 32 independent directions (b = 1000 s/mm^2) and six b = 0 images. Functional images were acquired using a T2*-sensitive gradient echo sequence (voxel size: 2 x 2 x 4 mm, TE = 30.0 ms, TR = 2000 ms, flip angle = 80°), and contained 37 slices of 3.4 mm thickness, with a 0.5 mm gap between slices. Volume acquisition occurred in an interleaved fashion. 330 volumes were obtained for each of the two functional runs in **study 1** and **2**, and 210 volumes were obtained for each of the three functional runs in **study 4**. During acquisition of fMRI time series, heart rate and respiratory data were collected using a breathing belt and a pulse oximeter.

fMRI analysis

fMRI analysis was performed in **study 1**, and in **study 4**. In house software (dicom2usb, http://dicom-port.com/) was used to de-identify all neuroimaging scans. Functional MRI analyses were performed in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). The preprocessing pipeline included slice-time correction, realignment, coregistration to the T1-weighted image, movement correction and normalization to MNI space. For normalization, we used a diffeomorphic registration algorithm (DARTEL (Ashburner 2007)) with spatial resolution after normalization 2 x 2 x 2 mm. Data were smoothed with a final Gaussian kernel equivalent to a standard 8 mm (see below). The fMRI time series data were high-pass filtered with a 128s cut-off, and whitened with an AR(1) model. For each participant, the canonical hemodynamic response function was used to compute their statistical model.

During the preprocessing of **study 1**, movement parameters produced by SPM's coregistration algorithm showed that 15 participants moved >3 mm in any direction during functional runs of the TAB, a common observation for older cohorts (Churchill et al., 2017). To correct for movement artifacts produced as a consequence of this, we used the ArtRepair toolbox (Mazaika et al. 2009; Levy and Wagner 2011) for **study 1**, **2**, and **4**. ArtRepair compares the amount of motion between volume acquisitions based on the mean intensity plot of all functional scans, and linearly interpolates scans in which motion exceeds a specified threshold. We used the recommended threshold value of 1.5% deviation from the mean intensity between scans. The average number of interpolated scans for our participants was 12.2 (1.8%) (SD = 19.6 (3.0%)) in **study 1**, and 14.0 (2.2%) (SD = 26.6 (4.2%)) in **study 4**. One participant was excluded from both fMRI tasks for showing movement >1.0 mm in $>25\%$ of scans, in line with ArtRepair's recommendations. ArtRepair smooths the individual subject data with a Gaussian smoothing kernel of 4 mm before normalization and movement correction. A Gaussian kernel of 7 mm was then used for the normalization to MNI space, resulting in a smoothed, normalized image equivalent to a standard 8 mm smoothed normalized image.

For all GLMs, SPM motion regressors were added to the design matrix as regressors of no interest. Additionally, 18 parameters correcting for physiological noise as recorded by a heartbeat detector and breathing belt during the scanning sessions. These were calculated using the PhysiO toolbox version r671 (https://www.tnu.ethz.ch/en/software/tapas.html).

First-level analyses are described in more detail under the individual studies. All second-level maps were produced with a family-wise error (FWE) voxel and cluster-corrected threshold at $p<0.05$, with a minimum cluster size of $k=10$.

Statistical analyses

In **study 1**, we used independent sample one-tailed t-tests to assess group differences in task performance, based on previously reported observations of impaired value-based decision-making performance in old age (Chowdhury et al., 2013a; Eppinger et al., 2011, 2015; Mell et al., 2005). Non-parametric independent two-tailed two sample Mann-Whitney tests were used to assess group differences in model parameters and other variables that were non-normally distributed. Two-tailed two-sample t-tests were used elsewhere. Pearson's correlations were used to analyse the data further, controlling for age and model fit, as defined by the participant's log likelihood, where appropriate. Statistical analyses were performed in SPSS.

In **study 2**, we used linear regression models, using the *lm* function in the R *stats* package. We predicted the strength of the value signal in vmPFC from 1) age, 2) dopamine D1-R availability in NAcc, and 3) fractional anisotropy in the tract between vmPFC-NAcc. Additionally, we predicted behavioural performance from the same three predictors, as well as 4) the strength of the value signal in vmPFC. Linear models were compared using the Bayesian Information Criterion (BIC).

In **study 3**, we investigated the interaction between action and valence with a 2 x 2 type III ANOVA using the *afex* package in R. The proportion of correct responses was the dependent variable, and action and valence independent within-subject variables with two levels (go/no-go and win/lose, respectively). To quantify the bias and study its relationship with D1-R availability measures, we calculated the overall action by valence interaction, reflecting the bias effect on the four conditions (GW + NGL – GL – NGW) and the bias effect on "win" conditions (GW – NGW). Two other behavioural measures reflecting the motivational bias were the instrumental bias parameter κ, and the overall performance on NGW. We then performed Pearson's correlation analysis between these behavioural scores and component scores of our PCA. We performed 10,000 permutations where we shuffled the values of the four measures of behavioral biases within participants and correlated these shuffled columns with the dopamine D1-R component loadings. The maximum t statistic from the four correlations in each iteration (four columns with shuffled values) was saved and added to the null distribution. This created null distributions that take into account the correlations between the measures of behavioral bias, and corrected for multiple comparisons taking this fact into account. Correlations in the data with an absolute t statistic that exceeded the absolute t statistic at the 95th percentile of this new null distribution were considered significant. Adjusted P values were calculated by counting the number of t values in the new null distribution that exceeded the observed t value, divided by 10,000.

In **study 4**, we used 2x2x2x3 MANOVAs to investigate the interactions between action, valence, group and dopamine D1-R availability. Parameter estimates were mean corrected and entered into the MANOVA. For significant interactions with dopamine components, individual bivariate correlations (and partial correlations controlling for age) were performed as follow-up analyses.

Exclusion and analysis samples for different studies

Not all 60 participants were included in all four studies.

In **study 1**, 3 participants were excluded from the fMRI sample. One was excluded due to not varying their choice behaviour on the TAB at all, one because of a malfunctioning button box, and one because of excessive head motion. The total final sample was 27 older participants and 30 younger participants (26 and 30 for PET).

In **study 2,** 12 additional participants were excluded due to the failure of white-matter tract reconstruction. The total final sample for this study was 23 older participants and 23 younger participants (22 and 23 for PET).

In **study 3**, those participants who performed badly on the task (as separated out by a 2-means clustering analysis based on GW performance) were initially excluded from behavioural analysis. In the main paper, 24 younger and 17 older participants were included (28 older and 30 younger for the PET analysis).

In **study 4**, bad performers were excluded from analysis. The final sample for this study included those participants whose performance exceeded 70% correct in each of the four conditions in the task, and were not the same participants as those reported in **study 3**. The final analyses include 28 younger and 17 older participants. This threshold was set based on what we judged to be good performance and is arbitrary.

For **study 1** and **study 2**, a number of participants were excluded for technical reasons. However, in **study 3** and **study 4**, a relatively large number of participants were excluded due to low performance on the valenced Go/No-Go task (see Task Descriptions). Exclusion was done before data analysis wherever possible, and as much as possible motivated by previous literature, and what we believed to be appropriate sample characteristics in light of the research question. In **study 3**, exclusion was done blindly by performing a two-mean clustering analysis. In **study 4**, I wanted to study action and inaction anticipation and processing in participants who were fully aware of task requirements. If task performance is very low, that is a strong indication that participants were not fully aware of task requirements, and that the participants are not anticipating what the fractal images instructed them to anticipate.

Statistical disclosure statement

Several iterations of participant selection and exclusion can contribute to "researcher degrees of freedom". This refers to the number of possible different variations of data collection and analyses that can be done as a result of the common practice of making decisions about data collection and analysis while research is ongoing. Intentional or unintentional questionable research practices can arise because of such decisions, which may only be followed through if the consequence of the decision yields statistical significance (Simmons et al., 2011). Participants were not excluded (and should never be excluded) because initial results did not match my hypotheses. However, I am aware that unconscious biases and justifications may still have influenced my decision to exclude certain participants, or choose certain analyses (Meehl, 1967). I have tried to counter this by showing analyses with and without covariates. In addition, I have reported the similarities or differences between important analyses in full and partial samples. In addition, below I provide a transparent summary of the analyses that were not reported in the articles or manuscripts, but that were performed. Although not perfect, this will allow the reader to more fully gauge the evidential value of the results, as preregistration reports are lacking.

In **study 1**, all fMRI analyses were initially performed without the motion correction step during the preprocessing stage. In addition, during the computational modelling analysis, I tried additional computational models that were not reported, of the "win-stay-lose-shift" family, which provided a worse fit to behaviour. I also attempted to decode stimulus identity from the orbitofrontal cortex using a support vector machine, which was unsuccessful. These analyses were not reported. Finally, I performed a GLM analysis on the fMRI data using neural activity related to switching behaviour and confidence. This was part of the initial manuscript, but was removed at the review stage. A copy of the reviewer comments and response to reviewers clarifying this decision is published together with the article on the eLife website.

In **study 2**, all of the performed analyses are reported.

In **study 3**, I initially attempted to fit a structural equation model to the PET data, after which I switched to the PCA that is reported in the paper. The results of the structural equation modelling are not reported.

In **study 4**, I also performed the fMRI analysis with temporal derivatives included in the GLM. I attempted a searchlight RSA analysis within the striatum, to try and find voxels that represented the action by valence interaction, which was unsuccessful. Additionally, I performed a hierarchical Bayesian regression analysis to estimate individual parameter estimates that reflected the extent to which people

represented anticipatory signals related to action, valence and the interaction between them in the different regions of interest I report in the current paper. I correlated these estimates with component scores of the PCA reported in **study 3** and **4**. Lastly, I performed a partial least squares analysis to obtain neural patterns of action and valence anticipation. None of these analyses yielded statistically significant or clearly interpretable results, and are not reported.

Throughout the years it took to complete the work in this thesis, I have increasingly realised the extent to which researcher degrees of freedom affect the quality and reproducibility of research and scientific reporting. Especially during the earlier works in this thesis, I may not have been as aware of the pitfalls of scientific enthusiasm (Gelman and Loken, 2013) as I am now. Although I stand behind the work and analyses done for this thesis, I believe that scientific practice in general and my own future research in particular could benefit from more narrowly specified and, ideally, preregistered analysis pipelines and hypotheses, as well as the publication of null results and (un)successful replication attempts.

Data and code availability

For **study 1**, all the data and code needed to reproduce the results are available at https://elifesciences.org/articles/26424.

For **study 2**, the code used to create the figures and the manuscript is available at https://github.com/liekelotte/DWI.

For **study 3**, the code for the computational modelling is available at https://github.com/liekelotte/GNG-models and the code and data for the PCA is posted at https://github.com/liekelotte/PCA-D1.

For **study 4**, analysis scripts will be made available upon submission of the manuscript.

INDIVIDUAL STUDIES

Study 1

Attenuation of dopamine-modulated prefrontal value signals underlies probabilistic reward learning deficits in old age

Study 1 was published in September 2017 in the journal *eLife,* with co-authors Jan Axelsson, Lars Nyberg, Peter Dayan, Katrine Riklund, Lars Bäckman and Marc Guitart-Masip.

Aim

Dopamine decline has been suggested to be part of the reason for worse probabilistic reward learning observed in older people, affecting neural processing in prefrontal cortex, striatum or both. Previous studies have found that RPE signals in the NAcc of older adults lack an expected-value component. Other brain activity important for reward processing and decision-making includes activity in vmPFC. To investigate which neural mechanism is affected by aging, we compared behaviour and brain activity of older participants during a probabilistic reward learning task that evokes RPEs to the behaviour and brain activity of younger participants performing the same task. To investigate the role of dopamine in the signalling of RPEs and value signals elsewhere in the brain, we also measured D1-R availability with PET.

Methods

We used a Two-Armed Bandit task (TAB), which participants performed in an fMRI scanner. The task required participants to pick one of two stimuli, and the payoff probability for each stimulus varied over the course of the experiment. We used computational modelling to quantify behaviour and generate expected value signals for the fMRI analysis. We looked at brain activity correlating with obtained reward and expected value, as well as the combination of the two, representing RPEs. We also used PET to measure dopamine D1-R availability. We estimated three first-level general linear models for fMRI. GLM 1 was set up to investigate how value anticipation is represented in the brain and how this representation differs between age groups. GLM 2 (with one parametric modulator for the putative RPE) and GLM 3 (with two parametric modulators (R and Q) for the canonical RPE) were set up to investigate the differences in the expression of the RPE signal at the time of outcome presentation in the old and young sample. After extracting parameter estimates from the relevant clusters, we investigated the relationships between these neural signals and dopamine D1-R availability with correlation analyses.

Results

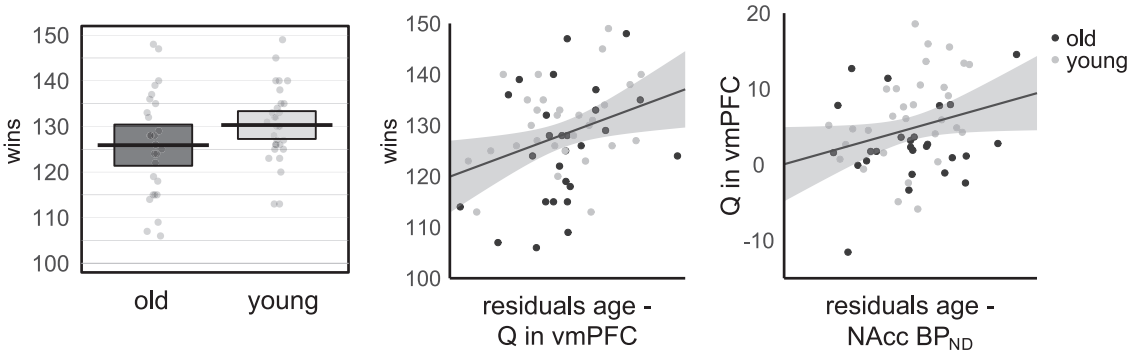

Figure 6. *left: difference in task performance between young and older participants. Younger participants earned more money on the TAB on average ($t(49) = 1.69$, p(one-tailed) $= 0.048$). middle: Behavioural performance was significantly predicted by the strength of the BOLD signal reflecting expected value (Q) in vmPFC ($r(53) = 0.37$, $p = 0.006$ when controlling for age and model fit). For display purposes, the correlations are shown with residuals after regressing out age. right: dopamine D1 BP_{ND} in NAcc is positively related to Q in vmPFC ($r(53) = 0.28$, $p = 0.038$, when controlling for age). For display purposes the correlations are shown with residuals after regressing out age.*

We found that anticipatory value signals in ventromedial prefrontal cortex (vmPFC) were attenuated in older adults. The strength of this signal predicted performance beyond age and was modulated by dopamine D1-R availability in NAcc. We did not observe a difference between age groups in the representation of RPEs.

Conclusion

These results uncover a value-anticipation mechanism in vmPFC that declines in aging, and that this mechanism is associated with dopamine D1-R availability in the NAcc. These results provide insights into the neural and behavioural underpinnings of probabilistic learning and highlight the mechanisms by which age-related dopaminergic deterioration impacts decision making.

Study 2

Corticostriatal white matter integrity and dopamine D1 receptor availability independently predict age differences in prefrontal value signaling during reward learning

Study 2 is currently under review. Co-authors are Benjamin Garzon, Jan Axelsson, Katrine Riklund, Lars Nyberg, Lars Backman and Marc Guitart-Masip.

Aim

Previous studies have found that the white matter tract that connects medial PFC with NAcc degrades in aging, and that this degradation can account for

the decline in probabilistic reward learning. This matches well with our finding from **study 1**, which showed that dopamine D1-R availability in the NAcc could predict the strength of value signals in vmPFC, which in turn predicted performance. Our aim for **study 2** was to investigate integrity of the white matter tract between NAcc and vmPFC using diffusion tensor imaging (DTI). The questions we answered is whether the integrity of these white matter tracts mediates the effect that the dopamine D1-R density in NAcc has on value representation in the vmPFC as described in study 1, or whether it independently contributes to an attenuation in vmPFC value signal. Additionally, we investigated the relationship between this white matter tract and performance on the TAB.

Methods

We used the same PET and fMRI data that we used in **study 1**. Participants performed the TAB in an fMRI scanner. The task required participants to pick one of two stimuli, and the payoff probability for each stimulus varied over the course of the experiment. We used the behavioural, computational and fMRI results from **study 1**. In addition, we performed deterministic tractography analysis, and calculated fractional anisotropy in the white matter tracts that connect vmPFC with NAcc. Then we predicted 1) the performance on the task from value signals in vmPFC, dopamine D1-R availability in the accumbens, and FA in the tract between accumbens and vmPFC, as well as 2) the strength of the value signal in vmPFC from age, dopamine D1-R availability in the NAcc and FA in the tract between accumbens and vmPFC. We expected that white matter integrity in the vmPFC-NAcc tract would predict performance beyond age and mediate the effects of D1-R availability in NAcc on vmPFC value anticipation.

Results

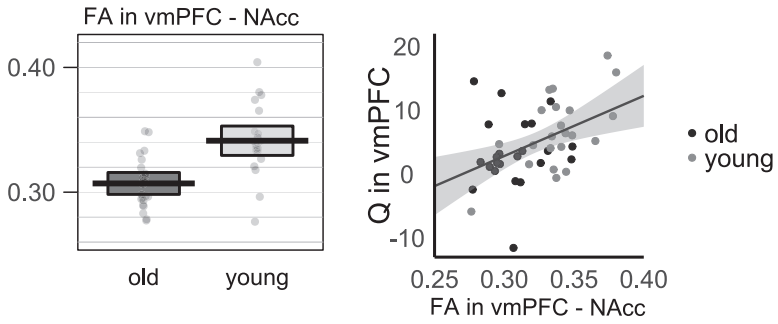

Figure 7. *left panel: We found that fractional anisotropy in the pathway between NAcc and vmPFC was significantly higher in younger compared to older participants ($M_{young}(SD) = 0.34\ (0.03)$, $M_{old}\ (SD) = 0.31\ (0.02)$, $t(40) = 5.00$, $p < 0.001$)). right panel: FA values in the pathway between NAcc and vmPFC were significantly correlated to the value-anticipatory activity in vmPFC ($r = 0.48$, $p = 0.001$). This correlation survived correction for age ($p = 0.01$).*

Table 1. Standardized coefficients and 95% confidence intervals predicting the expected-value signal in vmPFC. FA in the pathway between vmPFC and NAcc predicted the strength of the anticipatory value signal in vmPFC independently from age and from D1-R availability in NAcc.

Dependent: Q in vmPFC	β coefficient	95% confidence interval	p-value
age	0.30	-0.17 – 0.76	0.211
FA in vmPFC-NAcc	**0.48**	**0.12 – 0.69**	**0.006**
D1 BP_{ND} in NAcc	**0.41**	**-0.00 – 0.82**	**0.052**

Adjusted R^2 = 0.256, model p = 0.002.

Table 2. Standardized coefficients and 95% confidence intervals predicting performance. Only the expected-value signal in vmPFC significantly predicted performance on the TAB.

Dependent: wins	β coefficient	95% confidence interval	p-value
age	-0.33	-0.82 – 0.16	0.176
Q in vmPFC	**0.54**	**0.22 – 0.86**	**0.002**
FA in vmPFC-NAcc	-0.11	-0.49 – 0.27	0.550
D1 BP_{ND} in NAcc	-0.20	-0.64 – 0.24	0.373

Adjusted R^2=0.228, model p=0.006.

Conclusion

These results build on the result in **study 1**, showing that a value-anticipation mechanism in vmPFC is associated with dopamine D1-R availability in the NAcc and FA in the tract between vmPFC and NAcc beyond the effect of age. This adds to a mechanistic understanding of age-related differences in probabilistic reward learning ability.

Study 3

Dorsal striatal dopamine D1 receptor availability predicts an instrumental bias in action learning

Study 3 was published in January 2019 in the journal *Proceedings of the National Academy of Sciences (PNAS),* with co-authors Jan Axelsson, Lars Nyberg, Rumana Chowdhury, Raymond J Dolan, Katrine Riklund, Lars Backman and Marc Guitart-Masip.

Aim

Dopaminergic manipulation with L-DOPA has been shown to modulate a learning bias that favours the coupling of action and valence. The locus of this modulation has been a matter of debate. We investigated the locus of endogenous dopamine modulation on motivational biases during learning.

Methods

We used a valenced go/no-go task, which participants performed outside the scanner. The task elicits a motivational bias. We investigated the nature of this bias with computational modelling. We also used PET to measure dopamine D1-R availability and performed factor analysis on the binding potentials in a selection of four cortical (limbic cortex, frontal associative areas, inferior parietal lobe, motor cortex) and three striatal (NAcc, caudate, putamen) ROIs after correcting for age. We hypothesised that dopamine D1-R availability in the striatum would predict the extent to which individuals show a motivational bias in the valenced go/no-go task, and that cortical D1-R availability would be negatively related to this bias.

Results

Our PCA provided a three-component solution which demonstrates that cortical, dorsal striatal and ventral striatal dopamine D1-Rs provide separate, independent sources of variance in dopamine receptor availability as measured by [^{11}C]SCH23390. Table 3 displays the component scores for each ROI. We replicated this factor solution in an independent dataset. Using computational modelling, we characterized the motivational bias in our sample as of instrumental nature. Only the inter-individual variation in the dorsal striatal component was related to the extent to which individuals showed an instrumental learning bias (figure 8). We did not observe a negative relationship between inter-individual variation in the cortical component and the instrumental learning bias, nor did we observe any age effects of this bias.

Table 3. Component loadings of the three principal components recovered from the PCA with varimax rotation. Cortical ROI BP_{ND}s loaded on component 1, dorsal striatal ROI BP_{ND}s loaded on component 2, and the NAcc solely loaded on component 3.

	component 1: **cortical**	component 2: **dorsal striatal**	component 3: **ventral striatal**
Caudate	.39	**.88**	.18
Putamen	.32	**.85**	.34
Nucleus accumbens	.24	.26	**.93**
dlPFC/vlPFC: BA 9,44,45,46	**.80**	.48	.22
Limbic cortex: lateral/medial OFC	**.72**	.39	.45
Premotor cortex: BA 4,6	**.92**	.19	.17
Parietal cortex: IPL	**.79**	.46	.20

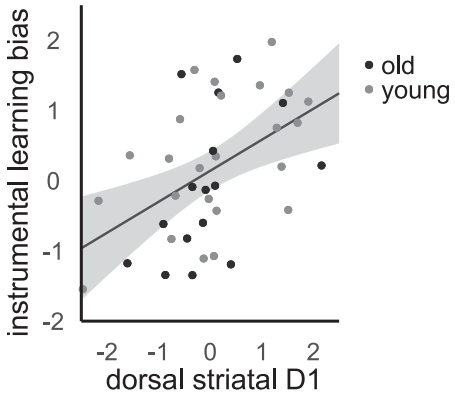

Figure 8. *Loadings on component 2 (dorsal striatal D1-R availability) correlated positively with the extent to which individuals displayed an instrumental learning bias in the Go/No-Go task ($r = 0.48$, $p = 0.005$).*

Conclusion

This suggests that D1-Rs are, to an extent, organised functionally in the human brain, and that this functional organisation affects behaviour. Specifically, dorsal striatal D1-R availability determines the extent to which individuals express an instrumental learning bias.

Study 4

The relationship between the representation of action and valence anticipatory patterns and dopamine D1 receptor availability

A first draft of this manuscript has been finalised with the co-authors Jan Axelsson, Katrine Riklund, Lars Nyberg, Lars Bäckman and Marc Guitart-Masip

Aim

Previous studies suggest that the representation of anticipated action dominates the representation of anticipated value in the striatum. Further, a previous study has shown that boosting the dopaminergic system with levodopa enhanced neural representations of actions, but only when these lead to rewarding outcomes. This has a bearing on the dopaminergic system and its integrity, as dopamine promotes actions coupled with valence. On the other hand, outcome processing shows a main effect of valence, with better than expected outcomes giving rise to activity in the NAcc, and worse than expected outcomes giving rise to activity in the insula, two central structures in the dopaminergic circuit. This study will investigate how endogenous dopamine integrity gives rise to patterns of action/valence representation, modulate outcome processing, and how these patterns are affected by aging.

Methods

We used the instructed go/no-go task, where participants learned the action/outcome contingencies and therefore performed at a high level. We predicted this would give rise to different neural patterns which will represent go and no-go patterns, possibly interacting with valence. Since levodopa enhanced the representation of anticipatory rewarding actions, and attenuated anticipatory rewarded inactions in a previous study, we expected to see enhanced GW activity (compared with NGW) in those with higher levels of dopamine D1 availability in striatum. In order to investigate this, we used 2x2x2x3 MANOVAs to investigate the interaction between dopamine receptor availability and neural representation of action and valence in the striatum. We also performed 2x2x2x3 MANOVAs on positive and negative RPE signalling to investigate outcome processing in the ventral striatum and insula.

Results

We found that in anticipatory activity, action dominated valence in all areas we investigated in the striatum. We did not observe a main effect of valence, nor did we observe an interaction between action and valence in anticipatory activity in the striatum. At outcome, the NAcc was responsive to better than expected outcomes, and the insula was responsive to worse than expected outcomes. We observed an attenuation of negative outcome processing in the insula in old age. The extent to which participants showed this attenuation was related to punishment sensitivity during the learning phase of the go/no-go task. In addition, we observed an interaction between cortical D1-R availability and differentiation between reward and punishment prediction errors in the insula (Figure 9).

Conclusion

Punishment processing in the insula may be attenuated in older adults, which may account for decreased punishment sensitivity in aging. Dopamine D1-R availability in the cortex may play a role in allowing punishment prediction errors in the insula to emerge.

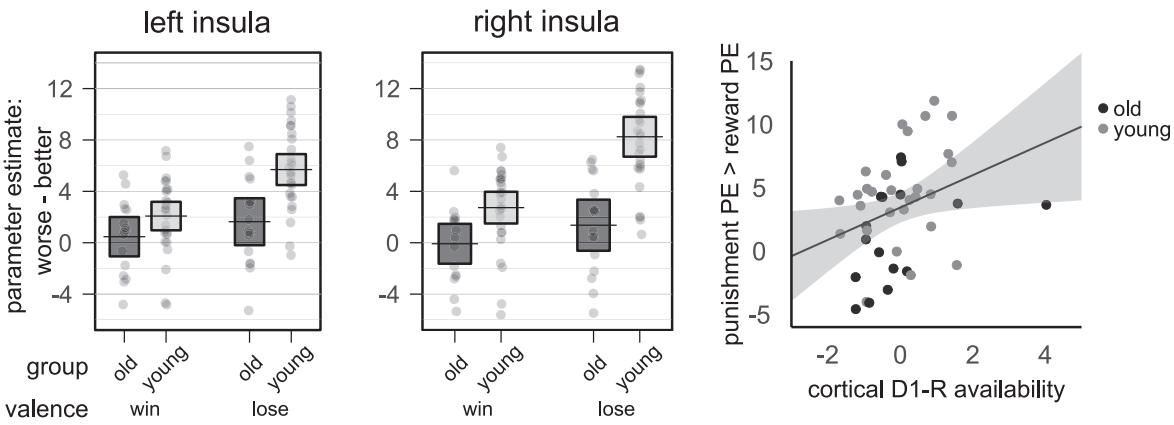

Figure 9. *We found a significant interaction between valence condition and age group in parameter estimates reflecting the difference between better and worse than expected outcomes in the insula in both hemispheres. The insula was more responsive to worse than expected outcomes than better than expected outcomes in younger participants (group x outcome interaction left insula: $F(1,39) = 13.50$, $p < 0.001$; right insula: $F(1,39) = 28.82$, $p < 0.001$). In addition, younger participants showed a stronger differentiation between punishment prediction errors and RPEs (group x valence interaction left insula: $F(1,39) = 5.24$, $p = 0.03$; right insula: $F(1,39) = 7.62$, $p = 0.009$). The three-way interaction reflecting enhanced differentiation for younger compared to older adults, in punishment conditions specifically was also significant in both hemispheres (group x outcome x valence left insula: $F(1,39) = 5.02$, $p = 0.03$; right insula: $F(1,39) = 10.59$, $p = 0.002$) right panel: We also found an interaction between outcome, valence condition and cortical D1-R availability, reflecting enhanced differentiation for those with higher dopamine D1-R component loading (cortical D1 x outcome x valence left insula: $F(1,39) = 6.92$, $p = 0.01$; right insula: $F(1,39) = 5.01$, $p = 0.03$. A subsequent correlation analysis revealed that this correlation was significant for the left insula (0.35, $p = 0.02$) and for the right insula ($r = 0.30$, $p = 0.05$). For display purposes, the averaged values between right and left insular activation are presented.*

DISCUSSION

In this thesis, I have investigated neural and behavioural correlates of value-based decision-making in a group of older and a group of younger adults. **Study 1 and 2** showed that value signalling in the vmPFC can predict performance on a probabilistic reward learning task. This value signal was attenuated in older, compared with younger adults. Two indicators of neural integrity predict the strength of this signal beyond age: First, the dopamine D1-R availability in the NAcc. Second, the integrity of the white matter connection between the NAcc and the vmPFC. We used a Bayesian observer model that moderates choices based on uncertainty and confidence measures to analyse behavioural data the TAB, and showed that this model fits behaviour better than a Rescorla-Wagner model that included perseveration and stickiness parameters.

Study 3 and **4** focused on dissociating the effects of action and valence on neural and behavioural correlates of decision-making. In **study 3**, we used computational modelling to characterize a commonly observed motivational bias as being a result of biased instrumental learning, as opposed to Pavlovian coupling of action and valence. **Study 3** also showed that we could decompose dopamine D1-R availability variability into cortical, dorsal striatal and ventral striatal components. Regardless of age, dopamine D1-R availability in the dorsal striatum was related to biased learning from rewarded actions. Lastly, **study 4** showed that once choice-outcome contingencies are learned, older adults did not differ from younger adults in anticipatory neural responses to action and valence. We did not observe a relationship between these anticipatory responses and dopamine D1-R availability. However, older adults showed an attenuated punishment prediction error signal in the insula. The strength of this punishment prediction error signal was related to dopamine D1-R availability in the cortex.

Reward prediction errors and striatal D1

One of the aims of this thesis was to investigate age differences between neural correlates of RPEs in the striatum. The TAB that was used in **study 1 and 2** is designed to continuously evoke behaviourally relevant RPEs, through the fluctuating probabilities of reward for each bandit arm. This requires participants to continuously update the relative values of each arm. In **study 4** on the other hand, we could only investigate RPEs that arose as a result of the task's probabilistic design, where occasional unexpected rewards or reward omissions violated what participants had learned, but did not instruct participants to update behaviour. This instructed nature of the task leads to stable estimates of the expected value component of RPEs for stimuli in **study 4**.

Due to the task design, we could investigate canonical RPE signals in the striatum in **study 1**. Canonical RPEs reflect the TD learning RPE at the time of reward delivery (Garrison et al., 2013; O'Doherty et al., 2003). Activity reflecting these RPEs should positively track the size of the reward delivered, but negatively track the expected value of the reward-predicting cue, as these two components both determine the size of the RPE (Behrens et al., 2008; Chowdhury et al., 2013a; Stenner et al., 2015). Chowdhury et al. (2013a) observed no expected value component of the RPE in older participants on placebo. When the dopaminergic system was boosted with L-DOPA however, this expected value component emerged (Chowdhury et al., 2013a). Because of the deterioration of the dopaminergic system with age, we expected that young people would show the same pattern of activity as older participants on L-DOPA in that study. Contrary to what we expected, we observed activity in the NAcc positively correlated with putative RPE (the simple difference between reward and expected value), but no activity in this area also negatively correlated with expected value in the young or older participants.

Because of the low temporal resolution of fMRI, it is perhaps not surprising that the two components of the RPE did not emerge in **study 1**. The BOLD signal may reflect a variety of distinct processes that are similarly energy-demanding, and therefore emerge as a seemingly unified signal (Haber and Behrens, 2014). Finer-grained methods like electrophysiology may be necessary to detect both RPE components (Stenner et al., 2015). The task design for **study 4** did not allow for an investigation into canonical prediction errors in the NAcc. Similarly to in **study 1**, we did observe a BOLD signal reflecting a putative RPE in **study 4** in the NAcc.

In neither of these studies did we observe a relationship between dopamine D1-R availability and putative RPEs in the NAcc. This is surprising in the light of evidence from animal and human studies. Namely, the hypothesis that the neural RPE would be related to dopamine D1-R availability rests on the assumption that dopaminergic bursts from the midbrain in response to differences between received and expected rewards are translated into a BOLD signal in the NAcc (Bayer and Glimcher, 2005; Soares-Cunha et al., 2016). D1-Rs have lower affinity to dopamine compared to D2-Rs, which implies that their activation requires phasic changes such as bursts of dopamine release in response to RPEs (Soares-Cunha et al., 2016). More availability of dopamine D1-Rs would result in more action potentials where D1 is expressed, which would in turn lead to more BOLD activation (Knutson and Gibbs, 2007). Boosting older adults' system with L-DOPA also amplified the expected value component of the neural RPE signal in the NAcc (Chowdhury et al., 2013a). In addition, studies in rats have shown D1-R blockade leads to an attenuation of NAcc activity in response to cocaine injections (Dixon et al., 2005; Marota et al., 2000).

However, the strength of activation in the NAcc is not only a result of the activation of postsynaptic D1-Rs. D2 autoreceptors on presynaptic terminals also regulate the amount of dopamine that is released in the NAcc, and agonizing these autoreceptors can attenuate BOLD responses in the NAcc (Chen et al., 2005). Thus, the relationship between D1-R availability and RPE is affected by the availability and activation of presynaptic D2 autoreceptors as well. Another possible explanation for the lack of relationship between the two can be found in the central position the NAcc has in the mesolimbic complex, receiving input from the midbrain, but also from a range of other cortical and subcortical areas (Grace et al., 2007). For example, a recent study in rats by Helbing et al. (2016) showed that stimulating the perforant pathway, which activates the mesolimbic system through the hippocampus, without causing dopamine release from the SN/VTA, resulted in BOLD activity in the NAcc. Importantly, this NAcc BOLD activation was still observed when rats were given the D1-R antagonist SCH23390, suggesting that BOLD activity in NAcc does not merely reflect D1-R activation. Conversely, the activation of the ACC/mPFC in rats as a result from perforant pathway stimulation was attenuated when rats were given SCH23390 (Helbing et al., 2016), suggesting that D1-R stimulation is necessary for the activation of prefrontal mesolimbic circuit components. This fits our finding in **study 1**, where we showed that D1-R availability in the NAcc was predictive of the BOLD signal reflecting the value signal in vmPFC, but not to RPEs.

Lastly, it is important to note that because the RPEs we observed were only putative, not canonical RPEs, they are highly correlated with the effect of reward alone. Since dopaminergic activity reflects the expected value component as well as the reward component of the RPE, it needs not to correlate with a neural signal reflecting reward alone.

Learning signals and striatal D1

Although we did not observe any relationship between D1-R availability in the NAcc and neural RPEs, we did find relationships between striatal dopamine D1-R availability and neural correlates of value-based decision-making throughout the brain across age groups. In **study 1**, this could be observed in the form of a relationship between NAcc D1-R availability and anticipatory value-related vmPFC activity during probabilistic reward learning. Further expanding on this finding, **study 2** demonstrated a relationship between the white-matter microstructural integrity of the pathway between NAcc and vmPFC and this anticipatory activity. In **study 3**, we showed a relationship between instrumental learning biases and dopamine D1-R availability in the dorsal striatum. All of these findings are in agreement with the theory of corticostriatal loops and the dopaminergic direct and indirect pathway modulate the activity and efficiency of these loops (Frank and O'Reilly, 2006; Haber and Behrens, 2014).

This theoretical framework posits that different cortical regions engage in different computations. These cortical regions all project to the striatum, in a functional topographic manner (see introduction), which converge there. Because of this functional organization, the striatum is thought to be central in the development of goal directed behaviour as a consequence of learning. Thus, good communication links between cortex and striatum, as well as an intact dopaminergic system are needed to make decisions that integrate these cortical computations.

Specifically, the vmPFC is known to provide one of the most robust value computation signals in fMRI studies, and lesions to this region in both hon-human primates and humans result in an inability to adapt value-based decisions to contingencies in the environment (Camille et al., 2011). Interestingly, a study where patients did not have to adapt their behaviour to a changing environment did not find a relationship between vmPFC lesion and behaviour, in this case on a monetary-incentive delay task (Pujara et al., 2016). This suggests that value-based decision-making and adapting behaviour to changing environmental contingencies may be specifically related to vmPFC function. Because of this robust signal, but also because of the consistent observation that damage to vmPFC impairs learning from rewards, the vmPFC is thought to be crucial for translating the valuation of one or more stimuli into a selection between stimuli (Hunt et al., 2012; Jocham et al., 2012).

The vmPFC has extensive projections to many different regions in the brain including the NAcc, and receives input from the NAcc as well, through the regulatory influence that NAcc activity has on basal ganglia and midbrain dopaminergic projections (Haber and Behrens, 2014). Specifically, D1-Rs in the NAcc may regulate the representation of rewarded choices in the vmPFC. Thus, it is possible that the D1-Rs in the NAcc play a role in gating the expected-value signal that emerges in the vmPFC, for example through downward projections from the vmPFC encouraging dopamine release in response to rewarding stimuli, which iteratively via the NAcc, arrives back in the vmPFC. Alternatively, dopamine release in response to highly valued stimuli may lead to an increased efficiency in communication between the NAcc and vmPFC in those individuals with high concentrations of dopamine D1-R availability. **Study 2** showed that the integrity of the white-matter microstructure between NAcc and vmPFC also positively correlated with the strength of the value-signal in vmPFC, supporting the premise that efficient direct communication between these two areas is crucial for the emergence of this value signal. Although this upregulation of preferred value representations can be explained with direct-indirect pathway circuitry, recent optogenetic findings have shown that 50% of dorsal ventral pallidum neurons, which are part of the indirect pathway, receive input from NAcc D1 MSNs as well as NAcc D2 MSNs. This suggests that the direct/indirect pathway architecture does not apply to NAcc projections (Kupchik et al., 2015). This may provide evidence for a more sophisticated role of D1 MSNs in the NAcc, balancing the direct and indirect pathway activity in this area.

Study 3 showed that dopamine D1-R availability in dorsal striatum (more specifically, in caudate and putamen) could predict the instrumental learning bias participants showed on a valenced go/no-go task where they had to learn action/valence contingencies. This bias, specifically, quantified the extent to which participants learned more from reward actions, compared to rewarded inactions. This finding fits the theory that the direct pathway, which D1-Rs are a part of, strengthens associations between rewards and the actions that lead to them. Specifically, activation of this pathway leads to the reinforcement of the previously executed action. The direct pathway becomes activated when dopamine neurons in the midbrain are active in response to unexpected rewards. Because the direct pathway then lowers the threshold for the repetition of the previously executed action (through increased connectivity between the thalamus and the cortex), this leads the direct pathway to couple actions with rewards, and discourages actions in response to punishments. The direct pathway activation leads to a lowering of the action threshold for actions that were previously rewarded (Kravitz et al., 2012). An increased availability in dopamine D1-Rs in dorsal striatum could therefore result in a more sensitive direct pathway, which encourages the actions in response to reward, but has more difficulty learning to put forward inaction patterns when rewarding stimuli are presented.

In summary, we have observed that dopamine D1-R availability in the striatum may be relevant in situations where individuals have to continuously update their representations of action propensities and reward. Based on previous evidence and a lack evidence for an association between performance or brain activity on the instructed Go/No-Go task after learning in the studies presented in this thesis, we could speculate that once choice-outcome associations are learned, as was the case in **study 4**, striatal D1-R availability do not show a clear relationship to neural or behavioural representations of expected value. Conversely, our data suggests that striatal dopamine D1-Rs are crucially important for the updating of cortical representations by gating corticostriatal loops.

Cortical D1-Rs and punishment processing in the insula

In an exploratory analysis, we found a relationship between individual loadings on the cortical D1 component and punishment prediction errors in the insula in **study 4**. The insula bears resemblance to the vmPFC in that it shows robust activation in response to value, but in the case of the insula, to negatively valued stimuli (Kim et al., 2006; Palminteri et al., 2012). It projects to the NAcc and likely also contributes to the updating of value representations that happens in the striatum (Haber and Behrens, 2014). Additionally, damage to the insula has shown to lead to impairments in punishment learning (Clark et al., 2008; Palminteri et al., 2012). The reason for a relationship between dopamine D1-R availability in the cortex and insular activation in response to punishment prediction errors could

stem from the presumed role of D1-Rs in the cortex in working memory (Seamans and Yang, 2004), where they are thought to increase the signal-to-noise ratio of working memory representations. Cortical D1-R availability may be required for the robust representations of punishment-related information in the insula. This, of course, rests on the assumption that the cortical D1 component we found in **study 3** reflects insular D1-R availability as well. As the insula was not included as a region of interest, this is an open question for future investigations.

An alternative explanation for the relationship between D1-R availability and neural activity in **study 4** can be found in the properties of the radiotracer that was used. [^{11}C]SCH23390 does not only bind to D1-Rs in the brain – it also shows a (albeit much lower) affinity for 5-HT$_{2A}$ receptors (Ekelund et al., 2007; Slifstein et al., 2007). This has been shown to affect binding potentials in the cortex. In the striatum, this is not as much of an issue because the number of D1-Rs is many times greater than 5-HT$_{2A}$ receptors. In the cortex however, 5-HT$_{2A}$ can represent up to 25% of the PET signal recorded with [^{11}C]SCH23390 (Ekelund et al., 2007). As our PCA was performed to separate the variability in D1-R binding into different components, I posit two possible interpretations of the first component. First, if cortical and striatal D1-R density in the brain share a large amount of variance, all of that variance could have been captured in component 2 or 3, leaving 5-HT$_{2A}$ variability to make up component 1. Second, if cortical and striatal D1-R density do not share a large amount of variance, component 1 could capture the variability in 5-HT$_{2A}$, as well as cortical D1-R availability, making this a component that at least describes the variability in both.

Serotonin has previously been theorized to be involved in punishment processing, and form an opponency axis with dopamine, although a two-dimensional spectrum to explain serotonin functioning has been acknowledged to be an oversimplification (Boureau and Dayan, 2011; Daw et al., 2002; den Ouden et al., 2013). Given the complexity of the serotonin system, which acts on 14 different receptor types (Boureau and Dayan, 2011), the exploratory nature of this finding, and the speculative nature of the involvement of the serotonin system, the possible serotonergic mechanisms at play will not be discussed further here.

Age effects on behavioural and neural correlates of decision-making

In line with the aims of this thesis, I have provided explanations for why older adults may be worse at value-based decision-making, compared to younger adults in three of the four studies considered (**study 1, 2 and 4**). In **study 1 and 2**, the observed relationship between neural correlates of expected value and performance on the task survived correction for age, suggesting that 1) this signal may get weaker as people age and 2) even within groups of individuals of the same

age, variability in the strength of this signal predicts performance. Note that the cross-sectional nature of these investigations does not allow drawing conclusions about the change in signal over time, but these observed relationships are the most convincing evidence one could observe in this dataset for how age-related neural changes may affect performance. Conversely, in **study 4**, we observed age-related attenuation of a punishment prediction error signal in the insula. This reduced neural signal was related to punishment sensitivity in learning during **study 3**, which was reduced in old age. However, the relationship between age-related attenuation and behaviour did not survive correction for age in this study. This could mean that neural integrity is affected by age, which consequently affects neural signalling and behaviour. It could also mean that age affects a range of different behavioural and neural factors that do not have an effect on each other.

It should be noted that in cases of age-mediation in cross sectional studies, mediator variables related to cross-sectional differences need not correlate with longitudinal change, and mediator variables unrelated to cross-sectional differences could correlate with longitudinal change (Lindenberger et al., 2011). This means that for **study 1 and 2**, longitudinal validation of our findings is required. Despite relatively robust evidence that longitudinal changes in white-matter microstructure do in fact occur, this cross-sectional mediation is no definite evidence for the existence of a causal change relationship. However, even though our data does not allow for causal reasoning, lesion studies support the behavioural relevance of the functional MRI and structural brain correlates that were correlated to age and relevant to behaviour in our sample, both for the vmPFC and the insula (Pujara et al., 2016; Weller et al., 2009).

It should also be noted that no age differences in instrumental learning bias parameter κ were observed in **study 3**, which is in line with findings from other cross-sectional datasets (Perosa et al., Betts et al., unpublished findings). Given a large difference between age groups in dopamine D1-R availability in the striatum, this suggests that D1-R availability only cannot predict the extent to which individuals show an instrumental learning bias, as that would result in a large behavioural difference between age groups, with the older group showing less instrumental learning bias compared to the younger group. Dopaminergic integrity is unlikely to be the only factor that drives reward learning behaviour, but it does seem to be able to explain a significant amount in the variability in an instrumental learning bias in two groups of different ages. However, D1-R availability in the striatum is not only factor that causes individuals to pair (or avoid pairing) action and valence. Other possible mechanisms for this observation could include D2-R availability (Richter et al., 2014), concentration of other catecholamines (Swart et al., 2017), dopamine synthesis capacity (Berry et al., 2016), and frontal executive control over such biases (Cavanagh et al., 2013).

In addition to the results discussed above, we also observed a slower learning rate in older adults, in **study 1** (Rescorla-Wagner model), and in **study 3** (both datasets). This is in line with previous findings showing that older adults generally adapt more slowly to quickly changing environments (Eppinger et al., 2011; Mell et al., 2005; Weiler et al., 2008). Similarly, in line with previous studies, we found a difference in sensitivity to punishments between age groups (Chowdhury et al., 2013b, 2014; Mata et al., 2011).

Interestingly, the evidence on attenuated responses to punishment prediction errors as we found in **study 4** in old age is mixed. Our result showed, specifically, that older adults' neural responses to punishment prediction errors (compared to negative RPEs during reward omission) were attenuated. Some previous studies have also found that outcome processing in the anterior insula is attenuated in older participants (Cox et al., 2008). Other studies show that outcome processing is similar between age groups in the insula (Samanez-Larkin et al., 2008), or that punishment anticipation, but not punishment outcome processing is attenuated in older adults (Samanez-Larkin et al., 2007). Similarly, although a range of studies found a decreased sensitivity to negative outcomes in older adults (Chowdhury et al., 2014; Harlé and Sanfey, 2012), some studies have also found an increased sensitivity to punishments in older adults, especially those above the age of 80 (Frank and Kong, 2008). On the other hand, a common observation in emotional processing and risk-taking in older adults is that they display an optimism bias (Berry et al., 2019; Kalenzaga et al., 2016; Nashiro et al., 2017; Reed and Carstensen, 2012) and are more willing to take risks in decision-making tasks where they learn from experience (Mata et al., 2011), which would suggest a reduced sensitivity to punishment. In addition, there is widespread evidence that the insula is one of the most atrophied brain areas in old age (Allen et al., 2005), which may account for its decreased activation during punishment prediction errors.

Computational modelling

In **study 1**, **2**, and **3** we used computational models that could simulate RL processes and Bayesian observer processes. In **study 1** and **2**, we showed that a Bayesian observer model described our observed behavioural data better compared to a more standard Rescorla-Wagner RL model. However, predictions from the Rescorla-Wagner model fit neural activity better in the NAcc, whereas predictions from the Bayesian observer model fit neural activity better in the vmPFC. This suggests that the processes that happen in different brain regions may best be described by different computational models.

Of course, the possible model space is infinite. Rescorla-Wagner and Bayesian observer model are only two types of models. Many more different types exist (Huys et al., 2016), and some different models could provide equally good

explanations for the same behavioural phenomena (Teufel and Fletcher, 2016). An example in the current thesis is the modelling procedure for the **study 3**. In this analysis, κ and π capture similar processes: whereas κ boosts learning from rewarded actions (and punished inactions), π boosts the value of an action when its potential outcome is a rewarded. These parameters tap into two different processes with similar manifestations in the current task. Although our model selection procedure indicated that a model with a single κ described our data best in **study 3**, data from a very similar dataset (with a slightly longer task runs) was best modelled with a model that also included π. In addition, κ in the datasets that we modelled sometimes took on negative values for a subset of participants, implying the opposite effect of a motivational bias. Investigations into this on the paradigm described here are currently ongoing, but it is important to emphasize that models and tasks should not be stretched beyond what reasonable neurobiological processes the model could capture, especially because other tasks have been developed to explicitly separate the instrumental component of this bias from the Pavlovian component (Swart et al., 2017). Modelling simulated performance data on the current task will shed light on the extent to which these models and their different processes can be disentangled, and to which extent they are reasonable in light of the underlying neurobiological hypotheses. The limits of reason should constantly be under scrutiny from the scientific community.

Limitations

In both versions of the valenced go/no-go task in this thesis, performance was relatively low compared to previous studies, especially in older adults (Chowdhury et al., 2013b). There may be several reasons for the relatively low performance on this task, but one obvious explanation is the distance between the responsible researchers on this project and those that collected data in Umeå. Because data was collected remotely, there was less hands-on researcher involvement during data collection. This means that we as experienced researchers who study behaviour on decision-making tasks were not present whenever instructions would have been misunderstood by participants or instructors. If this is indeed the reason, the impact of this distance on data quality has been significant. This highlights the importance of quality assurance and management by researchers on-site, especially for more complex experimental designs.

In **study 1, 2, and 4**, dopamine D1-R availability or microstructural white matter integrity was significantly correlated to neural signals that could predict performance, but not to performance itself. This indirect relationship to performance poses the question whether these measures of integrity in these cases really is important for good behaviour on these tasks. The direct relationship between dopamine D1-R availability and performance is perhaps more easily detected in

larger samples than the current one. In addition, one previous study demonstrated a relationship between white-matter microstructural integrity of this pathway and probabilistic reward learning (Samanez-Larkin et al., 2012). However, replications of the findings presented in this thesis are necessary to strengthen the validity of these findings.

Conclusion

The studies in this thesis provide important multimodal evidence that increases our understanding of the neural correlates that underlie value-based decision-making. First, we found a value signal in vmPFC that was affected by age, and could be predicted by dopaminergic integrity and microstructural integrity of white matter in the corticostriatal complex. Second, we demonstrated that D1-R availability in the striatum is related to the extent to which people learn from rewarded actions, compared to inactions. These studies have demonstrated that the existing theoretical framework surrounding the role of dopamine receptors in the basal ganglia fits with PET and behavioural data in older and younger adults. Importantly, these studies have also demonstrated that the dopaminergic system is complex, and that dopamine D1-R availability data alone cannot provide conclusive evidence on, for example, the role RPEs play in learning. In addition, other neurotransmitters and functional brain correlates likely play other important roles in decision-making in a complex environment.

ACKNOWLEDGEMENTS

I often wished I were more self-driven, but if I were, I may not have had so much fun working with others. There are a number of people who I owe thanks for their help and support throughout the years it took to complete the work in this thesis.

The person who deserves an endless amount of gratitude is my main supervisor **Marc Guitart-Masip**. Marc, we had such a good time! You are so good at what you do, you inspire others to be good as well. Thank you for being my friend, as well as my boss (and for that friendship to survive this PhD), and for running our lab with the perfect balance of freedom, focus and fun.

To my other supervisors, **Lars Bäckman**, thank you for your help, and for trusting me to do well. Your questions and friendly encouragements have been a perfect blend of challenging and motivating. **Lars Nyberg**, thank you for the input I received from you on manuscripts and on every occasion Marc and I visited Umeå. Your insights and comments have more than compensated for the physical distance between us while working together.

My closest friend and roommate during my PhD, **Nina Becker**, thanks for being the work spouse with whom I could (and did) share every thought. I genuinely hope we will work together in the future. Thanks also to my colleagues and friends who taught me about good science: **Pontus Plavén-Sigray**, **Granville Matheson**, **William Hedley Thompson**, and **Björn Schiffler**, thank you for inspiring me. I feel lucky to have had such gifted friends around during these past years.

My older science siblings who left the family home before me: **Rasmus Berggren**, you turned unproductive times into philosophical discussions that were more difficult than the work that we were supposed to do – thank you for exercising my cognitive reserve. **Ylva Köhncke**, thank you for your friendship – your ability to see things from multiple perspectives never ceases to amaze me. **Martin Bellander**, tack för att du tvingar mig att prata svenska ibland.

Thank you **Jan Axelsson**, for your PET expertise and your willingness to answer any question, no matter how basic. **Benjamín Garzón**, thanks for all the DTI help. Thanks to **Peter Dayan**, working with whom was a truly humbling experience. Thanks also to **Mats** and **Kajsa**, for their enormous efforts involved in data collection. Thanks to **Rumana Chowdhury**, **Ray Dolan** and **Anna Rieckmann** for sharing their data and for productive collaborations.

To the two excellent scientists who came and worked with Marc and me for short periods of time: **Emily Hird** and **Anni Richter**, thanks for the modelling discussions and afterwork sessions, and for making the most out of your time in Stockholm.

To everyone at ARC, the psychology, medical and social gerontology groups, and the administrative staff, thank you for your support. My roommates **Nic** and **Malin**, your concern for my plants is what has kept them alive for at least the past year, thank you! **Stina,** I appreciate always being allowed in (or is it never being kicked out of?) your office where you patiently listened to me rant. **Amaia**, thank you for always being there for me, and for being so genuinely empathic (que guay). Another colleague I am hugely indebted to in terms of support, especially towards the end of the thesis writing process, is **Bárbara**. Thank you for never tiring of helping me jump through bureaucratic hoops

Rita Almeida, you're a role model. Without you, our computational journal club wouldn't have kept everyone coming back. Thanks to everyone who attended that journal club who has helped me understand, among whom are **Ida**, **Philip**, **Irem**, **Nathalie, Gustav** and **Alexander**.

To my family, my brother **Aize**, and to both of my parents for being so (seemingly) totally okay with me just moving all over the place. Mam, **Kike**, thank you for making the beautiful cover, for trusting me and making sure I would not outrun myself. Thank you to my pap, **Albert**, for being interested and engaged whenever I would talk and think out loud about the nature of my work. Thanks to my grandparents **Anne, Aize, Ali and Synco** for demonstrating how best to get old. Tack, bedankt and thanks so much to the friends in Stockholm and elsewhere who put up with me during the more stressful periods of work and writing. You know who you are (and thanks for reading!).

Speaking of people who put up with me: **Gus,** you're the best – words will never be enough. Thanks for your cooking magic that somehow produces a casually fantastic dinner every day, and turns all obstacles into funny anecdotes.

REFERENCES

Aarts, E., Helmich, R.C., Janssen, M.J.R., Oyen, W.J.G., Bloem, B.R., and Cools, R. (2012). Aberrant reward processing in Parkinson's disease is associated with dopamine cell loss. NeuroImage *59*, 3339–3346.

Allen, J.S., Bruss, J., Brown, C.K., and Damasio, H. (2005). Normal neuroanatomical variation due to age: the major lobes and a parcellation of the temporal region. Neurobiol. Aging *26*, 1245–1260; discussion 1279-1282.

Arias-Carrión, O., Stamelou, M., Murillo-Rodríguez, E., Menéndez-González, M., and Pöppel, E. (2010). Dopaminergic reward system: a short integrative review. Int. Arch. Med. *3*, 24.

Asemi, A., Ramaseshan, K., Burgess, A., Diwadkar, V.A., and Bressler, S.L. (2015). Dorsal anterior cingulate cortex modulates supplementary motor area in coordinated unimanual motor behavior. Front. Hum. Neurosci. *9*, 309.

Bäckman, L., Nyberg, L., Lindenberger, U., Li, S.-C., and Farde, L. (2006). The correlative triad among aging, dopamine, and cognition: Current status and future prospects. Neurosci. Biobehav. Rev. *30*, 791–807.

Badre, D., Doll, B.B., Long, N.M., and Frank, M.J. (2012). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. Neuron *73*, 595–607.

Bartra, O., McGuire, J.T., and Kable, J.W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. NeuroImage *76*, 412–427.

Bayer, H.M., and Glimcher, P.W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron *47*, 129–141.

Bayer, H.M., Lau, B., and Glimcher, P.W. (2007). Statistics of midbrain dopamine neuron spike trains in the awake primate. J. Neurophysiol. *98*, 1428–1439.

Bear, M.F., Connors, B.W., and Paradiso, M.A. (2007). Neuroscience (Lippincott Williams & Wilkins).

Behrens, T.E.J., Woolrich, M.W., Walton, M.E., and Rushworth, M.F.S. (2007). Learning the value of information in an uncertain world. Nat. Neurosci. *10*, 1214–1221.

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., and Rushworth, M.F.S. (2008). Associative learning of social value. Nature *456*, 245–249.

Beierholm, U., Guitart-Masip, M., Economides, M., Chowdhury, R., Duzel, E., Dolan, R., and Dayan, P. (2013). Dopamine modulates reward related vigor. Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.

Berry, A.S., Shah, V.D., Baker, S.L., Vogel, J.W., O'Neil, J.P., Janabi, M., Schwimmer, H.D., Marks, S.M., and Jagust, W.J. (2016). Aging Affects Dopaminergic Neural Mechanisms of Cognitive Flexibility. J. Neurosci. *36*, 12559–12569.

Berry, A.S., Jagust, W.J., and Hsu, M. (2019). Age-related variability in decision-making: Insights from neurochemistry. Cogn. Affect. Behav. Neurosci. *19*, 415–434.

Bódi, N., Kéri, S., Nagy, H., Moustafa, A., Myers, C.E., Daw, N., Dibó, G., Takáts, A., Bereczki, D., and Gluck, M.A. (2009). Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. Brain *132*, 2385–2395.

Boorman, E.D., Behrens, T.E.J., Woolrich, M.W., and Rushworth, M.F.S. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. Neuron *62*, 733–743.

Boureau, Y.-L., and Dayan, P. (2011). Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol. *36*, 74–97.

Camille, N., Griffiths, C.A., Vo, K., Fellows, L.K., and Kable, J.W. (2011). Ventromedial Frontal Lobe Damage Disrupts Value Maximization in Humans. J. Neurosci. *31*, 7527–7532.

Carlsson, A. (1959). The occurrence, distribution and physiological role of catecholamines in the nervous system. Pharmacol. Rev. *11*, 490–493.

Cavanagh, J.F., Eisenberg, I., Guitart-Masip, M., Huys, Q., and Frank, M.J. (2013). Frontal theta overrides pavlovian learning biases. J. Neurosci. Off. J. Soc. Neurosci. *33*, 8541–8548.

Chao, L.L., and Knight, R.T. (1997). Prefrontal deficits in attention and inhibitory control with aging. Cereb. Cortex N. Y. N 1991 *7*, 63–69.

Chau, B.K.H., Kolling, N., Hunt, L.T., Walton, M.E., and Rushworth, M.F.S. (2014). A neural mechanism underlying failure of optimal choice with multiple alternatives. Nat. Neurosci. *17*, 463–470.

Chen, Y.-C.I., Choi, J.-K., Andersen, S.L., Rosen, B.R., and Jenkins, B.G. (2005). Mapping dopamine D2/D3 receptor function using pharmacological magnetic resonance imaging. Psychopharmacology (Berl.) *180*, 705–715.

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q., Düzel, E., and Dolan, R.J. (2013a). Dopamine restores reward prediction errors in old age. Nat. Neurosci. *16*, 648–653.

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dolan, R.J., and Düzel, E. (2013b). Structural integrity of the substantia nigra and subthalamic nucleus predicts flexibility of instrumental learning in older-age individuals. Neurobiol. Aging *34*, 2261–2270.

Chowdhury, R., Sharot, T., Wolfe, T., Düzel, E., and Dolan, R.J. (2014). Optimistic update bias increases in older age. Psychol. Med. *44*, 2003–2012.

Churchill, N.W., Raamana, P., Spring, R., and Strother, S.C. (2017). Optimizing fMRI preprocessing pipelines for block-design tasks as a function of age. NeuroImage *154*, 240–254.

Clark, D., and White, F.J. (1987). D1 dopamine receptor--the search for a function: a critical evaluation of the D1/D2 dopamine receptor classification and its functional implications. Synap. N. Y. N *1*, 347–388.

Clark, L., Bechara, A., Damasio, H., Aitken, M.R.F., Sahakian, B.J., and Robbins, T.W. (2008). Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making. Brain *131*, 1311–1322.

Clark, L., Studer, B., Bruss, J., Tranel, D., and Bechara, A. (2014). Damage to insula abolishes cognitive distortions during simulated gambling. Proc. Natl. Acad. Sci. U. S. A. *111*, 6098–6103.

Clatworthy, P.L., Lewis, S.J.G., Brichard, L., Hong, Y.T., Izquierdo, D., Clark, L., Cools, R., Aigbirhio, F.I., Baron, J.-C., Fryer, T.D., et al. (2009). Dopamine release in dissociable striatal subregions predicts the different effects of oral methylphenidate on reversal learning and spatial working memory. J. Neurosci. Off. J. Soc. Neurosci. *29*, 4690–4696.

Cohen, J.D., McClure, S.M., and Yu, A.J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *362*, 933–942.

Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature *482*, 85–88.

Collins, A.G.E., and Frank, M.J. (2014). Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol. Rev. *121*, 337–366.

Cools, R. (2008). Role of dopamine in the motivational and cognitive control of behavior. Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry *14*, 381–395.

Cools, R., and D'Esposito, M. (2011). Inverted-U-shaped dopamine actions on human working memory and cognitive control. Biol. Psychiatry *69*, e113-125.

Cools, R., Frank, M.J., Gibbs, S.E., Miyakawa, A., Jagust, W., and D'Esposito, M. (2009). Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. J. Neurosci. Off. J. Soc. Neurosci. *29*, 1538–1543.

Cools, R., Nakamura, K., and Daw, N.D. (2011). Serotonin and Dopamine: Unifying Affective, Activational, and Decision Functions. Neuropsychopharmacology *36*, 98–113.

Cox, K.M., Aizenstein, H.J., and Fiez, J.A. (2008). Striatal outcome processing in healthy aging. Cogn. Affect. Behav. Neurosci. *8*, 304–317.

Cox, S.M.L., Frank, M.J., Larcher, K., Fellows, L.K., Clark, C.A., Leyton, M., and Dagher, A. (2015). Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes. NeuroImage *109*, 95–101.

Croxson, P.L., Walton, M.E., O'Reilly, J.X., Behrens, T.E.J., and Rushworth, M.F.S. (2009). Effort-based cost-benefit valuation and the human brain. J. Neurosci. Off. J. Soc. Neurosci. *29*, 4531–4541.

Daunizeau, J., Ouden, H.E.M. den, Pessiglione, M., Kiebel, S.J., Stephan, K.E., and Friston, K.J. (2010). Observing the Observer (I): Meta-Bayesian Models of Learning and Decision-Making. PLOS ONE *5*, e15554.

Daw, N.D. (2011). Trial-by-trial data analysis using computational models: (Tutorial Review) - Oxford Scholarship. In Decision Making, Affect, and Learning, (Oxford, England: Oxford University Press), p.

Daw, N.D., and Doya, K. (2006). The computational neurobiology of learning and reward. Curr. Opin. Neurobiol. *16*, 199–204.

Daw, N.D., and Tobler, P.N. (2013). Neuroeconomics: Chapter 15. Value Learning through Reinforcement: The Basics of Dopamine and Reinforcement Learning (Elsevier Inc. Chapters).

Daw, N.D., Kakade, S., and Dayan, P. (2002). Opponent interactions between serotonin and dopamine. Neural Netw. Off. J. Int. Neural Netw. Soc. *15*, 603–616.

Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., and Dolan, R.J. (2006). Cortical substrates for exploratory decisions in humans. Nature *441*, 876–879.

Dayan, P., and Sejnowski, T.J. (1996). Exploration Bonuses and Dual Control. Mach. Learn. *25*, 5–22.

Dayan, P., Kakade, S., and Montague, P.R. (2000). Learning and selective attention. Nat. Neurosci. *3 Suppl*, 1218–1223.

Dixon, A.L., Prior, M., Morris, P.M., Shah, Y.B., Joseph, M.H., and Young, A.M.J. (2005). Dopamine antagonist modulation of amphetamine response as detected using pharmacological MRI. Neuropharmacology *48*, 236–245.

Ekelund, J., Slifstein, M., Narendran, R., Guillin, O., Belani, H., Guo, N.-N., Hwang, Y., Hwang, D.-R., Abi-Dargham, A., and Laruelle, M. (2007). In vivo DA D1 receptor selectivity of NNC 112 and SCH 23390. Mol. Imaging Biol. MIB Off. Publ. Acad. Mol. Imaging *9*, 117–125.

Eppinger, B., Hämmerer, D., and Li, S.-C. (2011). Neuromodulation of reward-based learning and decision making in human aging. Ann. N. Y. Acad. Sci. *1235*, 1–17.

Eppinger, B., Heekeren, H.R., and Li, S.-C. (2015). Age-related prefrontal impairments implicate deficient prediction of future reward in older adults. Neurobiol. Aging *36*, 2380–2390.

Farovik, A., Place, R.J., McKenzie, S., Porter, B., Munro, C.E., and Eichenbaum, H. (2015). Orbitofrontal cortex encodes memories within value-based schemas and represents contexts that guide memory retrieval. J. Neurosci. Off. J. Soc. Neurosci. *35*, 8333–8344.

Fiorillo, C.D. (2013). Two dimensions of value: dopamine neurons represent reward but not aversiveness. Science *341*, 546–549.

Fobbs, W.C., and Mizumori, S.J.Y. (2014). Cost-benefit decision circuitry: proposed modulatory role for acetylcholine. Prog. Mol. Biol. Transl. Sci. *122*, 233–261.

Frank, M.J., and Fossella, J.A. (2011). Neurogenetics and pharmacology of learning, motivation, and cognition. Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol. *36*, 133–152.

Frank, M.J., and Kong, L. (2008). Learning to avoid in older age. Psychol. Aging *23*, 392–398.

Frank, M.J., and O'Reilly, R.C. (2006). A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. Behav. Neurosci. *120*, 497–517.

Frank, M.J., Seeberger, L.C., and O'reilly, R.C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science *306*, 1940–1943.

Frank, M.J., Doll, B.B., Oas-Terpstra, J., and Moreno, F. (2009). The neurogenetics of exploration and exploitation: Prefrontal and striatal dopaminergic components. Nat. Neurosci. *12*, 1062–1068.

Garrison, J., Erdeniz, B., and Done, J. (2013). Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. Neurosci. Biobehav. Rev. *37*, 1297–1310.

Gazzaley, A., Cooney, J.W., Rissman, J., and D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. Nat. Neurosci. *8*, 1298–1300.

Gelman, A., and Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. 17.

Geurts, D.E.M., Huys, Q.J.M., den Ouden, H.E.M., and Cools, R. (2013). Serotonin and aversive Pavlovian control of instrumental behavior in humans. J. Neurosci. Off. J. Soc. Neurosci. *33*, 18932–18939.

Gorsuch, R.L. (2014). Factor Analysis: Classic Edition (Routledge).

Grace, A.A., Floresco, S.B., Goto, Y., and Lodge, D.J. (2007). Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. Trends Neurosci. *30*, 220–227.

Grady, C. (2012). The cognitive neuroscience of ageing. Nat. Rev. Neurosci. *13*, 491–505.

Gruber, A.J., Dayan, P., Gutkin, B.S., and Solla, S.A. (2006). Dopamine modulation in the basal ganglia locks the gate to working memory. J. Comput. Neurosci. *20*, 153.

Guitart-Masip, M., Fuentemilla, L., Bach, D.R., Huys, Q.J.M., Dayan, P., Dolan, R.J., and Duzel, E. (2011). Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. J. Neurosci. Off. J. Soc. Neurosci. *31*, 7867–7875.

Guitart-Masip, M., Huys, Q.J.M., Fuentemilla, L., Dayan, P., Duzel, E., and Dolan, R.J. (2012a). Go and no-go learning in reward and punishment: Interactions between affect and effect. NeuroImage *62*, 154–166.

Guitart-Masip, M., Chowdhury, R., Sharot, T., Dayan, P., Duzel, E., and Dolan, R.J. (2012b). Action controls dopaminergic enhancement of reward representations. Proc. Natl. Acad. Sci.

Guitart-Masip, M., Economides, M., Huys, Q.J.M., Frank, M.J., Chowdhury, R., Duzel, E., Dayan, P., and Dolan, R.J. (2014a). Differential, but not opponent, effects of l-DOPA and citalopram on action learning with reward and punishment. Psychopharmacology (Berl.) *231*, 955–966.

Guitart-Masip, M., Duzel, E., Dolan, R., and Dayan, P. (2014b). Action versus valence in decision making. TRENDS Cogn. Sci. *18*, 194–202.

Haber, S.N., and Behrens, T.E.J. (2014). The neural network underlying incentive-based learning: implications for interpreting circuit disruptions in psychiatric disorders. Neuron *83*, 1019–1039.

Haber, S.N., and Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol. *35*, 4–26.

Hagmann, P., Jonasson, L., Maeder, P., Thiran, J.-P., Wedeen, V.J., and Meuli, R. (2006). Understanding Diffusion MR Imaging Techniques: From Scalar Diffusion-weighted Imaging to Diffusion Tensor Imaging and Beyond. RadioGraphics *26*, S205–S223.

Halfmann, K., Hedgcock, W., Kable, J., and Denburg, N.L. (2016). Individual differences in the neural signature of subjective value among older adults. Soc. Cogn. Affect. Neurosci. *11*, 1111–1120.

Hall, H., Sedvall, G., Magnusson, O., Kopp, J., Halldin, C., and Farde, L. (1994). Distribution of D1- and D2-Dopamine Receptors, and Dopamine and Its Metabolites in the Human Brain. Publ. Online 01 Dec. 1994 Doi101038sjnpp1380111 *11*, 245–256.

Hamid, A.A., Pettibone, J.R., Mabrouk, O.S., Hetrick, V.L., Schmidt, R., Vander Weele, C.M., Kennedy, R.T., Aragona, B.J., and Berke, J.D. (2016). Mesolimbic Dopamine Signals the Value of Work. Nat. Neurosci. *19*, 117–126.

Harlé, K.M., and Sanfey, A.G. (2012). Social economic decision-making across the lifespan: An fMRI investigation. Neuropsychologia *50*, 1416–1424.

Hart, A.S., Rutledge, R.B., Glimcher, P.W., and Phillips, P.E.M. (2014). Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. J. Neurosci. Off. J. Soc. Neurosci. *34*, 698–704.

Helbing, C., Brocka, M., Scherf, T., Lippert, M.T., and Angenstein, F. (2016). The role of the mesolimbic dopamine system in the formation of blood-oxygen-level dependent responses in the medial prefrontal/anterior cingulate cortex during high-frequency stimulation of the rat perforant pathway. J. Cereb. Blood Flow Metab. *36*, 2177–2193.

Hikida, T., Kimura, K., Wada, N., Funabiki, K., and Nakanishi, S. (2010). Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. Neuron *66*, 896–907.

Howard, J.D., Gottfried, J.A., Tobler, P.N., and Kahnt, T. (2015). Identity-specific coding of future rewards in the human orbitofrontal cortex. Proc. Natl. Acad. Sci. U. S. A. *112*, 5195–5200.

Huang, A.C.W., Shyu, B.-C., and Hsiao, S. (2010). Dose-dependent dissociable effects of haloperidol on locomotion, appetitive responses, and consummatory behavior in water-deprived rats. Pharmacol. Biochem. Behav. *95*, 285–291.

Hunt, L.T., Kolling, N., Soltani, A., Woolrich, M.W., Rushworth, M.F.S., and Behrens, T.E.J. (2012). Mechanisms underlying cortical activity during value-guided choice. Nat. Neurosci. *15*, 470–476.

Hunt, L.T., Malalasekera, W.M.N., Berker, A.O. de, Miranda, B., Farmer, S.F., Behrens, T.E.J., and Kennerley, S.W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. Nat. Neurosci. *21*, 1471.

Huys, Q., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R.J., and Dayan, P. (2011). Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. PLoS Comput. Biol. *7*.

Huys, Q.J.M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., and Roiser, J.P. (2012). Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. PLoS Comput. Biol. *8*, e1002410.

Huys, Q.J.M., Maia, T.V., and Frank, M.J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. Nat. Neurosci. *19*, 404–413.

Hwang, E.J. (2013). The basal ganglia, the ideal machinery for the cost-benefit analysis of action plans. Front. Neural Circuits *7*.

Janowski, V., Camerer, C., and Rangel, A. (2013). Empathic choice involves vmPFC value signals that are modulated by social processing implemented in IPL. Soc. Cogn. Affect. Neurosci. *8*, 201–208.

Jocham, G., Hunt, L.T., Near, J., and Behrens, T.E.J. (2012). A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. Nat. Neurosci. *15*, 960–961.

Jocham, G., Brodersen, K.H., Constantinescu, A.O., Kahn, M.C., Ianni, A.M., Walton, M.E., Rushworth, M.F.S., and Behrens, T.E.J. (2016). Reward-Guided Learning with and without Causal Attribution. Neuron *90*, 177–190.

Kaasinen, V., Vilkman, H., Hietala, J., Någren, K., Helenius, H., Olsson, H., Farde, L., and Rinne, J. (2000). Age-related dopamine D2/D3 receptor loss in extrastriatal regions of the human brain. Neurobiol. Aging *21*, 683–688.

Kalenzaga, S., Lamidey, V., Ergis, A.-M., Clarys, D., and Piolino, P. (2016). The positivity bias in aging: Motivation or degradation? Emot. Wash. DC *16*, 602–610.

Keeler, J.F., Pretsell, D.O., and Robbins, T.W. (2014). Functional implications of dopamine D1 vs. D2 receptors: A "prepare and select" model of the striatal direct vs. indirect pathways. Neuroscience *282*, 156–175.

Kim, H., Shimojo, S., and O'Doherty, J.P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol. *4*, e233.

Kim, H., Shimojo, S., and O'Doherty, J.P. (2011). Overlapping Responses for the Expectation of Juice and Money Rewards in Human Ventromedial Prefrontal Cortex. Cereb. Cortex *21*, 769–776.

Klein-Flügge, M.C., Barron, H.C., Brodersen, K.H., Dolan, R.J., and Behrens, T.E.J. (2013). Segregated encoding of reward-identity and stimulus-reward associations in human orbitofrontal cortex. J. Neurosci. Off. J. Soc. Neurosci. *33*, 3202–3211.

Knutson, B., and Gibbs, S.E.B. (2007). Linking nucleus accumbens dopamine and blood oxygenation. Psychopharmacology (Berl.) *191*, 813–822.

Kobayakawa, M., Tsuruya, N., and Kawamura, M. (2010). Sensitivity to reward and punishment in Parkinson's disease: An analysis of behavioral patterns using a modified version of the Iowa gambling task. Parkinsonism Relat. Disord. *16*, 453–457.

Kravitz, A.V., Tye, L.D., and Kreitzer, A.C. (2012). Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat. Neurosci. *15*, 816.

Kringelbach, M.L., and Rolls, E.T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Prog. Neurobiol. *72*, 341–372.

Kupchik, Y.M., Brown, R.M., Heinsbroek, J.A., Lobo, M.K., Schwartz, D.J., and Kalivas, P.W. (2015). Coding the direct/indirect pathways by D1 and D2 receptors is not valid for accumbens projections. Nat. Neurosci. *18*, 1230–1232.

Laskowski, C.S., Williams, R.J., Martens, K.M., Gruber, A.J., Fisher, K.G., and Euston, D.R. (2016). The role of the medial prefrontal cortex in updating reward value and avoiding perseveration. Behav. Brain Res. *306*, 52–63.

Leisman, G., Melillo, R., and Carrick, F.R. (2013). Clinical Motor and Cognitive Neurobehavioral Relationships in the Basal Ganglia. Basal Ganglia - Integr. View.

Lindenberger, U., and Baltes, P.B. (1997). Intellectual Functioning in Old and Very Old Age: Cross-Sectional Results From the Berlin Aging Study. Psychol. Aging *12*, 410–432.

Lindenberger, U., von Oertzen, T., Ghisletta, P., and Hertzog, C. (2011). Cross-sectional age variance extraction: what's change got to do with it? Psychol. Aging *26*, 34–47.

Marcott, P.F., Mamaligas, A.A., and Ford, C.P. (2014). Phasic Dopamine Release Drives Rapid Activation of Striatal D2-Receptors. Neuron *84*, 164–176.

Marota, J.J., Mandeville, J.B., Weisskoff, R.M., Moskowitz, M.A., Rosen, B.R., and Kosofsky, B.E. (2000). Cocaine activation discriminates dopaminergic projections by temporal response: an fMRI study in Rat. NeuroImage *11*, 13–23.

Marsden, C.A. (2006). Dopamine: the rewarding years. Br. J. Pharmacol. *147*, S136–S144.

Mata, R., Josef, A.K., Samanez-Larkin, G.R., and Hertwig, R. (2011). Age differences in risky choice: a meta-analysis. Ann. N. Y. Acad. Sci. *1235*, 18–29.

Meehl, P.E. (1967). Theory-Testing in Psychology and Physics: A Methodological Paradox. Philos. Sci. *34*, 103–115.

Mell, T., Heekeren, H.R., Marschner, A., Wartenburger, I., Villringer, A., and Reischies, F.M. (2005). Effect of aging on stimulus-reward association learning. Neuropsychologia *43*, 554–563.

Montague, P.R., Dolan, R.J., Friston, K.J., and Dayan, P. (2012). Computational psychiatry. Trends Cogn. Sci. *16*, 72–80.

Moscufo, N., Wakefield, D.B., Meier, D.S., Cavallari, M., Guttmann, C.R.G., White, W.B., and Wolfson, L. (2018). Longitudinal microstructural changes of cerebral white matter and their association with mobility performance in older persons. PLOS ONE *13*, e0194051.

Moustafa, A.A., Cohen, M.X., Sherman, S.J., and Frank, M.J. (2008). A role for dopamine in temporal decision making and reward maximization in parkinsonism. J. Neurosci. Off. J. Soc. Neurosci. *28*, 12294–12304.

Nashiro, K., Sakaki, M., Braskie, M.N., and Mather, M. (2017). Resting-state networks associated with cognitive processing show more age-related decline than those associated with emotional processing. Neurobiol. Aging *54*, 152–162.

Nieuwenhuis, S., Aston-Jones, G., and Cohen, J.D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. Psychol. Bull. *131*, 510–532.

Niv, Y., Edlund, J.A., Dayan, P., and O'Doherty, J.P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. J. Neurosci. Off. J. Soc. Neurosci. *32*, 551–562.

Noonan, M.P., Walton, M.E., Behrens, T.E.J., Sallet, J., Buckley, M.J., and Rushworth, M.F.S. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. Proc. Natl. Acad. Sci. *107*, 20547–20552.

O'Doherty, J.P. (2007). Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. Ann. N. Y. Acad. Sci. *1121*, 254–272.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (2003). Temporal Difference Models and Reward-Related Learning in the Human Brain. Neuron *38*, 329–337.

Ogawa, S., Lee, T.M., Kay, A.R., and Tank, D.W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. U. S. A. *87*, 9868–9872.

Old, S.R., and Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: a meta-analysis. Psychol. Aging *23*, 104–118.

Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., and Lorenzetti, V. (2018). The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. Hum. Brain Mapp. *39*, 3398–3418.

den Ouden, H.E.M., Daw, N.D., Fernandez, G., Elshout, J.A., Rijpkema, M., Hoogman, M., Franke, B., and Cools, R. (2013). Dissociable effects of dopamine and serotonin on reversal learning. Neuron *80*, 1090–1100.

Padoa-Schioppa, C., and Assad, J.A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Nat. Neurosci. *11*, 95–102.

Pagnoni, G., Zink, C.F., Montague, P.R., and Berns, G.S. (2002). Activity in human ventral striatum locked to errors of reward prediction. Nat. Neurosci. *5*, 97–98.

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., Durr, A., et al. (2012). Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. Neuron *76*, 998–1009.

Pan, W.-X., Schmidt, R., Wickens, J.R., and Hyland, B.I. (2005). Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. J. Neurosci. Off. J. Soc. Neurosci. *25*, 6235–6242.

Payzan-LeNestour, E., and Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. PLoS Comput. Biol. *7*, e1001048.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., and Frith, C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature *442*, 1042–1045.

Philiastides, M.G., Biele, G., and Heekeren, H.R. (2010). A mechanistic account of value computation in the human brain. Proc. Natl. Acad. Sci. *107*, 9430–9435.

Picciotto, M.R., Higley, M.J., and Mineur, Y.S. (2012). Acetylcholine as a Neuromodulator: Cholinergic Signaling Shapes Nervous System Function and Behavior. Neuron *76*, 116–129.

Preuschoff, K., 't Hart, B.M., and Einhauser, W. (2011). Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. Front. Neurosci. *5*.

Pujara, M.S., Philippi, C.L., Motzkin, J.C., Baskaya, M.K., and Koenigs, M. (2016). Ventromedial Prefrontal Cortex Damage Is Associated with Decreased Ventral Striatum Volume and Response to Reward. J. Neurosci. Off. J. Soc. Neurosci. *36*, 5047–5054.

Raja Beharelle, A., Polanía, R., Hare, T.A., and Ruff, C.C. (2015). Transcranial Stimulation over Frontopolar Cortex Elucidates the Choice Attributes and Neural Mechanisms Used to Resolve Exploration-Exploitation Trade-Offs. J. Neurosci. Off. J. Soc. Neurosci. *35*, 14544–14556.

Raz, N., Gunning-Dixon, F., Head, D., Rodrigue, K.M., Williamson, A., and Acker, J.D. (2004). Aging, sexual dimorphism, and hemispheric asymmetry of the cerebral cortex: replicability of regional differences in volume. Neurobiol. Aging *25*, 377–396.

Reed, A.E., and Carstensen, L.L. (2012). The Theory Behind the Age-Related Positivity Effect. Front. Psychol. *3*.

Rescorla, R.A., and Wagner, A.R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. In Classical Conditioning II: Current Research and Theory, (New York: Appleton-Century-Croft), p. 18.

Richter, A., Guitart-Masip, M., Barman, A., Libeau, C., Behnisch, G., Czerney, S., Schanze, D., Assmann, A., Klein, M., Düzel, E., et al. (2014). Valenced action/inhibition learning in humans is modulated by a genetic variant linked to dopamine D2 receptor expression. Front. Syst. Neurosci. *8*.

Rieckmann, A., Karlsson, S., Karlsson, P., Brehmer, Y., Fischer, H., Farde, L., Nyberg, L., and Bäckman, L. (2011). Dopamine D1 receptor associations within and between dopaminergic pathways in younger and elderly adults: links to cognitive performance. Cereb. Cortex N. Y. N 1991 *21*, 2023–2032.

Robertson, C.L., Ishibashi, K., Mandelkern, M.A., Brown, A.K., Ghahremani, D.G., Sabb, F., Bilder, R., Cannon, T., Borg, J., and London, E.D. (2015). Striatal D1- and D2-type Dopamine Receptors Are Linked to Motor Response Inhibition in Human Subjects. J. Neurosci. *35*, 5990–5997.

Rogers, R.D. (2011). The Roles of Dopamine and Serotonin in Decision Making: Evidence from Pharmacological Experiments in Humans. Neuropsychopharmacology *36*, 114–132.

Rönnlund, M., Nyberg, L., Bäckman, L., and Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. Psychol. Aging *20*, 3–18.

Ross, S., and Stearns, C. (2010). SharpIR: White paper.

Rouault, M., Drugowitsch, J., and Koechlin, E. (2019). Prefrontal mechanisms combining rewards and beliefs in human decision-making. Nat. Commun. *10*, 301.

Rudebeck, P.H., and Murray, E.A. (2014). The orbitofrontal oracle: cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. Neuron *84*, 1143–1156.

Rudebeck, P.H., Walton, M.E., Smyth, A.N., Bannerman, D.M., and Rushworth, M.F.S. (2006). Separate neural pathways process different decision costs. Nat. Neurosci. *9*, 1161–1168.

Rudorf, S., Preuschoff, K., and Weber, B. (2012). Neural Correlates of Anticipation Risk Reflect Risk Preferences. J. Neurosci. *32*, 16683–16692.

Rushworth, M.F.S., Noonan, M.P., Boorman, E.D., Walton, M.E., and Behrens, T.E. (2011). Frontal cortex and reward-guided learning and decision-making. Neuron *70*, 1054–1069.

Rutledge, R.B., Lazzaro, S.C., Lau, B., Myers, C.E., Gluck, M.A., and Glimcher, P.W. (2009). Dopaminergic Drugs Modulate Learning Rates and Perseveration in Parkinson's Patients in a Dynamic Foraging Task. J. Neurosci. *29*, 15104–15114.

Salthouse, T.A. (1992). Influence of processing speed on adult age differences in working memory. Acta Psychol. (Amst.) *79*, 155–170.

Samanez-Larkin, G.R., and Knutson, B. (2015). Decision making in the ageing brain: changes in affective and motivational circuits. Nat. Rev. Neurosci. *16*, 278–289.

Samanez-Larkin, G.R., Gibbs, S.E.B., Khanna, K., Nielsen, L., Carstensen, L.L., and Knutson, B. (2007). Anticipation of monetary gain but not loss in healthy older adults. Nat. Neurosci. *10*, 787–791.

Samanez-Larkin, G.R., Hollon, N.G., Carstensen, L.L., and Knutson, B. (2008). Individual Differences in Insular Sensitivity During Loss Anticipation Predict Avoidance Learning. Psychol. Sci. *19*, 320–323.

Samanez-Larkin, G.R., Levens, S.M., Perry, L.M., Dougherty, R.F., and Knutson, B. (2012). Frontostriatal White Matter Integrity Mediates Adult Age Differences in Probabilistic Reward Learning. J. Neurosci. Off. J. Soc. Neurosci. *32*, 5333–5337.

Samanez-Larkin, G.R., Worthy, D.A., Mata, R., McClure, S.M., and Knutson, B. (2014). Adult age differences in frontostriatal representation of prediction error but not reward outcome. Cogn. Affect. Behav. Neurosci. *14*, 672–682.

Sawaguchi, T., and Goldman-Rakic, P.S. (1991). D1 dopamine receptors in prefrontal cortex: involvement in working memory. Science *251*, 947–950.

Schoenfeld, R., Foreman, N., and Leplow, B. (2014). Ageing and spatial reversal learning in humans: findings from a virtual water maze. Behav. Brain Res. *270*, 47–55.

Schuck, N.W., Cai, M.B., Wilson, R.C., and Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. Neuron *91*, 1402–1412.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. J. Neurophysiol. *80*, 1–27.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Seamans, J.K., and Yang, C.R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. Prog. Neurobiol. *74*, 1–58.

Shokoueinejad, M., Park, D.-W., Jung, Y.H., Brodnick, S.K., Novello, J., Dingle, A., Swanson, K.I., Baek, D.-H., Suminski, A.J., Lake, W.B., et al. (2019). Progress in the Field of Micro-Electrocorticography. Micromachines *10*.

Simmons, J.P., Nelson, L.D., and Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychol. Sci. *22*, 1359–1366.

Singer, T., Critchley, H.D., and Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. Trends Cogn. Sci. *13*, 334–340.

Slifstein, M., Kegeles, L.S., Gonzales, R., Frankle, W.G., Xu, X., Laruelle, M., and Abi-Dargham, A. (2007). [11C]NNC 112 Selectivity for Dopamine D1 and Serotonin 5-HT2A Receptors: A PET Study in Healthy Human Subjects. J. Cereb. Blood Flow Metab. *27*, 1733–1741.

Smith, R.E., Tournier, J.-D., Calamante, F., and Connelly, A. (2012). Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. NeuroImage *62*, 1924–1938.

Smith, R.E., Tournier, J.-D., Calamante, F., and Connelly, A. (2015). SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. NeuroImage *119*, 338–351.

Soares-Cunha, C., Coimbra, B., Sousa, N., and Rodrigues, A.J. (2016). Reappraising striatal D1- and D2-neurons in reward and aversion. Neurosci. Biobehav. Rev. *68*, 370–386.

Stalnaker, T.A., Cooch, N.K., and Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. Nat. Neurosci. *18*, 620–627.

Stenner, M.-P., Rutledge, R.B., Zaehle, T., Schmitt, F.C., Kopitzki, K., Kowski, A.B., Voges, J., Heinze, H.-J., and Dolan, R.J. (2015). No unified reward prediction error in local field potentials from the human nucleus accumbens: evidence from epilepsy patients. J. Neurophysiol. *114*, 781–792.

Suhara, T., Fukuda, H., Inoue, O., Itoh, T., Suzuki, K., Yamasaki, T., and Tateno, Y. (1991). Age-related changes in human D1 dopamine receptors measured by positron emission tomography. Psychopharmacology (Berl.) *103*, 41–45.

Surmeier, D.J., Carrillo-Reid, L., and Bargas, J. (2011). Dopaminergic modulation of striatal neurons, circuits, and assemblies. Neuroscience *198*, 3–18.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement learning: an introduction (Cambridge, Mass.: MIT Press).

Swart, J.C., Froböse, M.I., Cook, J.L., Geurts, D.E., Frank, M.J., Cools, R., and den Ouden, H.E. (2017). Catecholaminergic challenge uncovers distinct Pavlovian and instrumental mechanisms of motivated (in)action. ELife *6*.

Takahashi, H., Yamada, M., and Suhara, T. (2012). Functional Significance of Central D1 Receptors in Cognition: Beyond Working Memory. J. Cereb. Blood Flow Metab. *32*, 1248–1258.

Teufel, C., and Fletcher, P.C. (2016). The promises and pitfalls of applying computational models to neurological and psychiatric disorders. Brain J. Neurol. *139*, 2600–2608.

Tobler, P.N., Fiorillo, C.D., and Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. Science *307*, 1642–1645.

Tournier, J.-D., Calamante, F., and Connelly, A. (2012). MRtrix: Diffusion tractography in crossing fiber regions. Int. J. Imaging Syst. Technol. *22*, 53–66.

Vijver, I. van de, Ridderinkhof, K.R., Harsay, H., Reneman, L., Cavanagh, J.F., Buitenweg, J.I.V., and Cohen, M.X. (2016). Frontostriatal anatomical connections predict age- and difficulty-related differences in reinforcement learning. Neurobiol. Aging *46*, 1–12.

Volkow, N.D., Fowler, J.S., Wang, G.J., Logan, J., Schlyer, D., MacGregor, R., Hitzemann, R., and Wolf, A.P. (1994). Decreased dopamine transporters with age in healthy human subjects. Ann. Neurol. *36*, 237–239.

Volkow, N.D., Wang, G.J., Fowler, J.S., Ding, Y.S., Gur, R.C., Gatley, J., Logan, J., Moberg, P.J., Hitzemann, R., Smith, G., et al. (1998). Parallel loss of presynaptic and postsynaptic dopamine markers in normal aging. Ann. Neurol. *44*, 143–147.

Von Siebenthal, Z., Boucher, O., Rouleau, I., Lassonde, M., Lepore, F., and Nguyen, D.K. (2017). Decision-making impairments following insular and medial temporal lobe resection for drug-resistant epilepsy. Soc. Cogn. Affect. Neurosci. *12*, 128–137.

Walton, M.E., Behrens, T.E.J., Buckley, M.J., Rudebeck, P.H., and Rushworth, M.F.S. (2010). Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron *65*, 927–939.

Wang, Y., and Ruhe, G. (2007). The Cognitive Process of Decision Making. Int. J. Cogn. Inform. Nat. Intell. IJCINI *1*, 73–85.

Weiler, J.A., Bellebaum, C., and Daum, I. (2008). Aging affects acquisition and reversal of reward-based associative learning. Learn. Mem. *15*, 190–197.

Weller, J.A., Levin, I.P., Shiv, B., and Bechara, A. (2009). The effects of insula damage on decision-making for risky gains and losses. Soc. Neurosci. *4*, 347–358.

West, R. (1999). Visual distraction, working memory, and aging. Mem. Cognit. *27*, 1064–1072.

Wickens, J. (1990). Striatal dopamine in motor activation and reward-mediated learning: steps towards a unifying model. J. Neural Transm. Gen. Sect. *80*, 9–31.

Wickens, J.R., Reynolds, J.N.J., and Hyland, B.I. (2003). Neural mechanisms of reward-related motor learning. Curr. Opin. Neurobiol. *13*, 685–690.

Wilson, R.C., Geana, A., White, J.M., Ludvig, E.A., and Cohen, J.D. (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. J. Exp. Psychol. Gen. *143*, 2074–2081.

Wimmer, G.E., Braun, E.K., Daw, N.D., and Shohamy, D. (2014). Episodic Memory Encoding Interferes with Reward Learning and Decreases Striatal Prediction Errors. J. Neurosci. *34*, 14901–14912.

Wu, M., Chang, L.C., Walker, L., Lemaitre, H., Barnett, A.S., Marenco, S., and Pierpaoli, C. (2008). Comparison of EPI distortion correction methods in diffusion tensor MRI using a novel framework. Med. Image Comput. Comput.-Assist. Interv. MICCAI Int. Conf. Med. Image Comput. Comput.-Assist. Interv. *11*, 321–329.